AD_____

Award Number: DAMD17-98-1-8469

TITLE: Collection of Prostate Cancer Families and Mapping Additional Hereditary Prostate Cancer Genes (HPC2, HPC3,…)

PRINCIPAL INVESTIGATOR: William B. Isaacs, Ph.D.

CONTRACTING ORGANIZATION: Johns Hopkins University
School of Medicine
Baltimore, Maryland 21205-2196

REPORT DATE: April 2003

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	April 2003	Final (1 Oct 1998 - 31 Mar 2003)

4. TITLE AND SUBTITLE
Collection of Prostate Cancer Families and Mapping Additional Hereditary Prostate Cancer Genes (HPC2, HPC3,...)

5. FUNDING NUMBERS
DAMD17-98-1-8469

6. AUTHOR(S)
William B. Isaacs, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Johns Hopkins University
School of Medicine
Baltimore, Maryland 21205-2196

E-Mail: wisaacs@jhmi.edu

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
Original contains color plates: All DTIC reproductions will be in black and white.

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 Words)
Our initial genome wide search for linkage in multiplex prostate cancer families implicated *1q24-25* as harboring a major prostate cancer susceptibility gene (*HPC1*), although there was significant evidence for locus heterogeneity and at least 5 other loci were implicated as the sites of HPC genes. Using families ascertained in Phase I of this project, an additional HPC locus located at *Xq27-28 (HPCX)* was identified. Furthermore we obtained strong preliminary evidence of an additional novel HPC locus on chromosome 8 that may account for up to 20% of all HPC families. To continue and extend these analyses towards the identification of the HPC genes on chromosomes X and 8, we proposed to narrow the gene-containing regions on Xq and 8p (from ~20 cM to 1-2 cM) using a variety of approaches including association studies in family- and population based sample, and to assess candidate genes in regions delineated by these approaches by mutation screening. Our main focus has been on chromosome 8 where a number of candidate genes have been evaluated. Importantly, one gene, *MSR1* at 8p22, was found to harbor clearly inactivating mutations, and other cancer-associated variants in HPC families and non-HPC cases. These results implicate for the first time genetic variation affecting macrophage function as an important determinant of inherited susceptibility for prostate cancer.

14. SUBJECT TERMS
Prostate cancer

15. NUMBER OF PAGES
121

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

Table of Contents

Introduction

In spite of the magnitude of the problem which prostate cancer presents, our understanding of the molecular mechanisms underlying prostatic carcinogenesis remains elusive. It is clear from the recent progress made in colorectal, renal and breast cancer that analysis of familial forms of common human neoplasms can yield unprecedented insight into the specific genetic mechanisms responsible for both hereditary and sporadic forms of such cancers. Segregation analyses supporting the existence of dominantly acting susceptibility genes have provided the basis for linkage studies in high-risk prostate cancer families. However, such studies are complicated by a number of factors, including genetic heterogeneity, late age of disease onset, and a high phenocopy rate due to the high disease prevalence and the lack of identified distinguishing features of hereditary prostate cancer (HPC) as compared to the sporadic form of the disease. Our initial genome wide search for linkage in multiplex prostate cancer families implicated *1q24-25* as harboring a major prostate cancer susceptibility gene (*HPC1*), although there was significant evidence for locus heterogeneity and at least 5 other loci were implicated as the sites of HPC genes (Smith et al 1996). In Phase I of this project, we proposed collection of additional, highly informative prostate cancer families and use of these families to search for additional HPC loci. We have made significant progress in these areas, and have now collected DNA samples from 159 HPC families. These families provided the basis for the identification of a novel HPC locus located at *Xq27-28 (HPCX)* (Xu et al 1998) Furthermore, in following up other genomic regions as part of Phase I, we have obtained strong preliminary evidence of an additional novel HPC locus on chromosome 8 that may account for up to 20% of all HPC families, and intriguingly, a potentially much higher proportion of Ashkenazi Jewish HPC families. To continue and extend these analyses towards the identification of the HPC genes on chromosomes X and 8, we proposed the following specific aims for Phase II: 1) Narrow the gene-containing regions on Xq and 8p (from ~20 cM to 1-2 cM) using the following approaches:; and 2) Assess candidate genes in regions delineated in Specific Aim 1. The MSR1 gene at 8p22 has emerged as an important candidate prostate cancer susceptibility gene from these studies (Xu et al 2002, Xu et al 2003), and provides important new insight into the mechanisms responsible for genetic susceptibility for prostate cancer.

Body

Statement of Work: 1& 2) Narrow the gene-containing regions on Xq and 8p, and assess candidate genes in regions delineated in Specific Aim 1:

Multiple prostate cancer linkage regions and candidate genes have been systematically evaluated in prostate cancer families on chromosomes 1, 3, 8 and X using haplotype analyses and conventional linkage and association analyses. Important findings, including both positive and negative results, have been obtained and published. These results significantly advance the knowledge in this field. They are summarized as the following:

1) Polymorphic GGC repeats in the androgen receptor gene are associated with hereditary and sporadic prostate cancer risk (Chang et al. 2002, Hum Genet 110:122-9). This study examined the AR gene, on the X chromosome.

Abstract: Androgen receptor (AR) has long been hypothesized to play an important role in prostate cancer etiology. Two trinucleotide repeat polymorphisms (CAG and GGC repeats in exon 1 of the AR gene) have been investigated as risk factors for prostate cancer in several studies. However, the results are inconclusive, probably because of the variations of study designs, characteristics of study samples, and choices of analytical methods. In this study, we evaluated evidence for linkage and association between the two AR repeats and prostate cancer by using the following comprehensive approaches: (1) a combination of linkage and association studies, (2) a test for linkage by parametric analysis and the male-limited X-linked transmission/disequilibrium test (XLRC-TDT), (3) a test for association by using both population-based and family-based tests, and (4) a study of both hereditary and sporadic cases. A positive but weak linkage score (HLOD=0.49, P=0.12) was identified in the AR region by parametric analysis; however, stronger evidence for linkage in the region, especially at the GGC locus, was observed in the subset of families whose proband had 16 GGC repeats (HLOD=0.70, P=0.07) or by using XLRC-TDT (z'=2.65, P=0.008). Significantly increased frequencies of the 16 GGC repeat alleles in 159 independent hereditary cases (71%) and 245 sporadic cases (68%) cases compared with 211 controls (59%) suggested that GGC repeats were associated with prostate cancer (P=0.02). Evidence for the association between the 16 GGC repeats and prostate cancer risk was stronger with XLRC-TDT (z'=2.66, P=0.007). No evidence for association between the CAG repeats and prostate cancer risk was observed. The consistent results from both linkage and association studies strongly implicate the GGC repeats in the AR as a prostate cancer susceptibility gene. Further studies on this polymorphism in other independent data sets and functional analysis of the GGC repeat length on AR activity are warranted.

2) Germline mutations in the ribonuclease L gene in families showing linkage with HPC1 (Carpten et al. 2002, Nature Genet 30:181-4). This study examined the RNASEL gene on chromosome 1.

Abstract: Although prostate cancer is the most common non-cutaneous malignancy diagnosed in men in the United States, little is known about inherited factors that influence its genetic predisposition. Here we report that germline mutations in the gene encoding 2'-5'-oligoadenylate (2-5A)-dependent RNase L (RNASEL) segregate in prostate cancer families that show linkage to the HPC1 (hereditary prostate cancer 1) region at 1q24-25 (ref. 9). We identified RNASEL by a

positional cloning/candidate gene method, and show that a nonsense mutation and a mutation in an initiation codon of RNASEL segregate independently in two HPC1-linked families. Inactive RNASEL alleles are present at a low frequency in the general population. RNASEL regulates cell proliferation and apoptosis through the interferon-regulated 2-5A pathway and has been suggested to be a candidate tumor suppressor gene. We found that microdissected tumors with a germline mutation showed loss of heterozygosity and loss of RNase L protein, and that RNASEL activity was reduced in lymphoblasts from heterozyogous individuals compared with family members who were homozygous with respect to the wildtype allele. Thus, germline mutations in RNASEL may be of diagnostic value, and the 2-5A pathway might provide opportunities for developing therapies for those with prostate cancer.

3) Joint effect of HSD3B1 and HSD3B2 genes is associated with hereditary and sporadic prostate cancer susceptibility (Chang et al. 2002, Cancer Res. 62:1784-9). This study examined the HSD3B genes on chromosome 1.

Abstract: 3beta-hydroxysteroid dehydrogenases (HSD3Bs), encoded by the HSD3B gene family at 1p13, have long been hypothesized to have a major role in prostate cancer susceptibility. The recent reports of a prostate cancer linkage at 1p13 provided additional evidence that HSD3B genes may be prostate cancer susceptibility genes. To evaluate the possible role of HSD3B genes in prostate cancer, we screened a panel of DNA samples collected from 96 men with or without prostate cancer for sequence variants in the putative promoter region, exons, exon-intron junctions, and 3'-untranslated region of HSD3B1 and HSD3B2 genes by direct sequencing. Eleven single nucleotide polymorphisms (SNPs) were identified, four of which, including a missense change (B1-N367T), were informative. These four SNPs were further genotyped in a total of 159 hereditary prostate cancer probands, 245 sporadic prostate cancer cases, and 222 unaffected controls. Although a weak association between prostate cancer risk and a missense SNP (B1-N367T) was found, stronger evidence for association was found when the joint effect of the two genes was considered. Men with the variant genotypes at either B1-N367T or B2-c7519g had a significantly higher risk to develop prostate cancer, especially the hereditary type of prostate cancer. Most importantly, the subset of hereditary prostate cancer probands, whose families provided evidence for linkage at 1p13, predominantly contributed to the observed association. Additional studies are warranted to confirm these findings.

4) Associations between hOGG1 sequence variants and prostate cancer susceptibility. (Xu et al. 2002, Cancer Res. 62:2253-7). This study examined the hOGG1 gene on chromosome 3.

Abstract: 8-Hydroxyguanine is a mutagenic base lesion produced by reactive oxygen species. The hOGG1 gene encodes a DNA glycosylase/AP lyase that can suppress the mutagenic effects of 8-hydroxyguanine by catalyzing its removal from oxidized DNA. A population-based (245 cases and 222 controls) and family-based (159 hereditary prostate cancer families) association study was performed to test the hypothesis that sequence variants of hOGG1 increase susceptibility to prostate cancer. We found that the genotype frequency of two sequence variants (11657A/G and Ser326Cys) was significantly different between cases and controls. The association with 11657A/G is confirmed and strengthened by our family-based association study. These results suggest that sequence variants in this gene are associated with prostate cancer risk, presumably through defective DNA repair function of hOGG1.

5) Evaluation of DLC1 as a prostate cancer susceptibility gene: mutation screen and association study (Zheng et al. 2003, Mutation Res, in press). This study examined the DLC1 gene on chromosome 8.

Abstract: A gene or genes on chromosome 8p22-23 have been implicated in prostate carcinogenesis by the observation of frequent deletions of this region in prostate cancer cells. More recently, two genetic linkage studies in hereditary prostate cancer (HPC) families suggest that germline variation in a gene in this region may influence prostate cancer susceptibility as well. DLC1 (deleted in liver cancer), a gene in this interval, has been proposed as a candidate tumor suppressor gene because of its homology (86% similarity) with rat p122 RhoGAP, which catalyzes the conversion of active GTP-bound rho complex to the inactive GDP-bound form, and thus suppresses Ras-mediated oncogenic transformation. A missense mutation and three intronic insertions/deletions in 126 primary colorectal tumors have been previously identified. However, there are no reports of DLC1 mutation screening in prostate tumors or in germ line DNA of prostate cancer patients. In this study, we report the results of the first mutation screen and association study of DLC1 in genomic DNA samples from hereditary and sporadic prostate cancer patients. The PCR products in the 5' UTR, all 14 exons, exon-intron junctions, and 3' UTR were directly sequenced in 159 HPC probands. Eight exonic nucleotide polymorphisms (SNPs) were identified, only one of which resulted in an amino acid change. Twenty-three other SNPs were identified in intronic regions. Seven informative SNPs that spanned the complete DLC1 gene were genotyped in an additional 249 sporadic cases and 222 unaffected controls. No significant difference in the allele and genotype frequencies were observed among HPC probands, sporadic cases, and unaffected controls. These results suggest that DLC1 is unlikely to play an important role in prostate cancer susceptibility.

6) Germline mutations and sequence variants of the macrophage scavenger receptor 1 gene are associated with prostate cancer risk. (Xu et al Nat Genet 2002 Oct;32(2):321-5). This study examines the MSR1 gene on chromosome 8.

Abstract: Deletions on human chromosome 8p22-23 in prostate cancer cells and linkage studies in families affected with hereditary prostate cancer (HPC) have implicated this region in the development of prostate cancer. The macrophage scavenger receptor 1 gene (MSR1, also known as SR-A) is located at 8p22 and functions in several processes proposed to be relevant to prostate carcinogenesis. Here we report the results of genetic analyses that indicate that mutations in MSR1 may be associated with risk of prostate cancer. Among families affected with HPC, we identified six rare missense mutations and one nonsense mutation in MSR1. A family-based linkage and association test indicated that these mutations co-segregate with prostate cancer ($P = 0.0007$). In addition, among men of European descent, MSR1 mutations were detected in 4.4% of individuals affected with non-HPC as compared with 0.8% of unaffected men ($P = 0.009$). Among African American men, these values were 12.5% and 1.8%, respectively ($P = 0.01$). These results show that MSR1 may be important in susceptibility to prostate cancer in men of both African American and European descent.

7) Germline Sequence Variants of the LZTS1 Gene are Associated with Prostate Cancer Risk (Hawkins et al. 2002, Cancer Genet Cytogenet. 2002 Aug;137(1):1-7). This study examined the LZTS1 gene on chromosome 8.

Abstract: The 8p22-23 region has been identified as a potential site for gene(s) associated with prostate cancer. The gene LZTS1 has been mapped to the 8p22-23 region and identified as a potential tumor suppressor based on LOH studies using primary esophageal tumors. Sequence analysis of mRNA from various tumors has revealed multiple mutations and aberrant mRNA transcripts. The most recent report associates LZTS1 function with stabilization of $p34^{cdc2}$ during the late S-G_2/M stage of mitosis, affecting normal cell growth. In this study, a detailed DNA sequence analysis of LZTS1 was performed in a screening panel consisting of sporadic and hereditary prostate cancer cases and unaffected controls. Twenty-four SNPs, 15 of which were novel, were identified in germline DNA. Four coding SNPs were identified. Eleven informative SNPs were genotyped in 159 HPC probands, 245 sporadic prostate cancer cases, and 222 unaffected controls. Four of these SNPs were statistically significant for association with prostate cancer ($p \leq 0.04$). A bioinformatic analysis of LZTS1was also performed, identifying additional features in the peptide sequence. These putative analyses suggest that the LZTS1 peptide plays a structural role in cell function. All of these results add evidence supporting a role of LZTS1 in prostate cancer risk.

8) Common sequence variants of the macrophage scavenger receptor 1 gene are associated with prostate cancer risk. (Xu et al 2003, Am J Hum Genet. 2003 Jan;72(1):208-12). This study examines the association of common germline variants in the MSR1 gene and prostate cancer risk.

Abstract: Rare germline mutations of macrophage scavenger receptor 1 (MSR1) gene were reported to be associated with prostate cancer risk in families with hereditary prostate cancer (HPC) and in patients with non-HPC (Xu et al. 2002). To further evaluate the role of MSR1 in prostate cancer susceptibility, at Johns Hopkins Hospital, we studied five common variants of MSR1 in 301 patients with non-HPC who underwent prostate cancer treatment and in 250 control subjects who participated in prostate cancer-screening programs and had normal digital rectal examination and PSA levels (<4 ng/ml). Significantly different allele frequencies between case subjects and control subjects were observed for each of the five variants (P value range.01-.04). Haplotype analyses provided consistent findings, with a significant difference in the haplotype frequencies from a global score test (P=.01). Because the haplotype that is associated with the increased risk for prostate cancer did not harbor any of the known rare mutations, it appears that the observed association of common variants and prostate cancer risk are independent of the effect of the known rare mutations. These results consistently suggest that MSR1 may play an important role in prostate carcinogenesis.

Key research accomplishments

We have evaluated eight genes as prostate cancer susceptibility genes, including the RNASEL, HSD3B1, and HSD3B2 on chromosome 1, the AR gene on the X chromosome, the hOGG1 gene on chromosome 3, and the DLC1, LZTS and MSR1 genes on chromosome 8. The finding of multiple mutations in this latter gene, including clearly inactivating mutations, in both familial and non familial prostate cancer, strongly implicates this gene as an important prostate cancer susceptibility gene. The fact that this gene is involved in the innate immune system suggests for the first time that genetic variation is the host response to endogenous and/or exogenous antigens may play a critical role in determining inherited risk for prostate cancer.

Reportable outcomes

Manuscripts published during the time period of April 1, 2001 and March 31, 2003:

Carpten J, Nupponen N, Isaacs S, Sood R, Robbins C, Xu J, Faruque M, Moses T, Ewing C, Gillanders E, Hu P, Bujnovszky P, Makalowska I, Baffoe-Bonnie A, Faith D, Smith J, Stephan D, Wiley K, Brownstein M, Gildea D, Kelly B, Jenkins R, Hostetter G, Matikainen M, Schleutker J, Klinger K, Connors T, Xiang Y, Wang Z, Demarzo A, Papadopoulos N, Kallioniemi OP, Burk R, Meyers D, Gronberg H, Meltzer P, Silverman R, Bailey-Wilson J, Walsh P, Isaacs W, Trent J. Germline mutations in the ribonuclease L gene in families showing linkage with HPC1. Nature Genetics 2002; 30:181-184

Chang B, Zheng SL, Isaacs SD, Wiley K, Carpten JD, Hawkins GA, Bleecker ER, Walsh PC, Trent JM, Meyers DA, Isaacs WB, Xu J (2001) Linkage and association of CYP17 gene in hereditary and sporadic prostate cancer. Int J Cancer 95: 354-359

Chang B, Zheng SL, Hawkins GA, Isaacs SD, Wiley KE, Turner A, Carpten JD, Bleecker ER, Walsh PC, Trent JM, Meyers DA, Isaacs WB, Xu J (2002) Joint effect of HSD3B1 and HSD3B2 genes is associated with hereditary and sporadic prostate cancer susceptibility. Cancer Res 62:1784-1789.

Chang B, Zheng SL, Hawkins GA, Isaacs SD, Wiley KE, Turner A, Carpten JD, Bleecker ER, Walsh PC, Trent JM, Meyers DA, Isaacs WB, Xu J (2002) Polymorphic GGC repeats in the androgen receptor gene are associated with hereditary and sporadic prostate cancer risk. Hum Genet 2002; 110:122-129.

Hawkins GA, Mychaleckyj JC, Zheng SL, Faith DA, Kelly B, Isaacs SD, Wiley KE, Chang BL, Ewing CM, Bujnovszky P, Bleecker ER, Walsh PC, Meyers DA, Isaacs WB, Xu J (2002). Germline Sequence Variants of the LZTS1 Gene are Associated with Prostate Cancer Risk. Cancer Genet Cytogenet (In Press).

Ho G, Knapp M, Freije D, Nelson WG, Smith JR, Carpten JD, Bailey-Wilson JE, Beaty TH, Petersen G, Xu J, Kamensky V, Walsh PC, Isaacs WB (2002) Linkage Analyses of Androgen Receptor and Glutathione S-Transferase Pi Variants in Prostate Cancer Families. International Journal of Cancer Int J Cancer 98:938-42.

Xu J, Zheng SL, Chang B, Isaacs SD, Wiley K, Hawkin GA, Bleecker ER, Walsh PC, Trent JM, Meyers DA, Isaacs WB (2001) Linkage and association studies of prostate cancer susceptibility gene on 8p22-23. Am J Hum Genet 69:341-350.

Xu J, Zheng SL, Turner A, Isaacs SD, Wiley K, Hawkins GA, Chang B, Bleecker ER, Walsh PC, Meyers DA, Isaacs WB (2002) Associations between hOGG1 sequence variants and prostate cancer susceptibility. Cancer Res 62:2253-2257.

Xu J, Zheng SL, Komiya A, Mychaleckyj JC, Isaacs SD, Hu JJ, Sterling D, Lange EM, Hawkins GA, Turner A, Ewing CM, Faith DA, Johnson JR, Suzuki H, Bujnovszky P, Wiley KE, DeMarzo AM, Bova GS, Chang B, Hall MC, McCullough DL, Partin AW, Kassabian VS, Carpten JD, Bailey-Wilson JE, Trent JM, Ohar J, Bleecker ER, Walsh PC, Isaacs WB, Meyers DA. (2002) Germline mutations and

sequence variants of the macrophage scavenger receptor 1 gene are associated with prostate cancer risk. Nat Genet 32:321-325.

Xu J, Zheng SL, Komiya A, Mychaleckyj JC, Isaacs SD, Chang B, Turner AR, Ewing CM, Wiley KE, Hawkins GA, Bleecker ER, Walsh PC, Meyers DA, Isaacs WB. Common sequence variants of the macrophage scavenger receptor 1 gene are associated with prostate cancer risk. (2003) Am J Hum Genet 72:208-212.

Zheng SL, Xu J, Chang B, Isaacs SD, Wiley K, Bleecker ER, Walsh PC, Trent JM, Meyers DA, Isaacs WB (2001) Evaluation of linkage of HPC20 in 159 hereditary prostate cancer pedigrees. Hum Genet 108:430-435.

Zheng SL, Mychaleckyj JC, Hawkins GA, Isaacs SD, Wiley KE, Turner A, Chang B, von Kap-Herr C, Carpten JD, Pettenati M, Bleecker ER, Walsh PC, Trent JM, Meyers DA, Isaacs WB, Xu J. Evaluation of DLC1 as a prostate cancer susceptibility gene: mutation screen and association study. Mut Res 2003 (In Press).

Conclusions

We have evaluated seven genes as prostate cancer susceptibility genes, including the RNASEL, HSD3B1, and HSD3B2 on chromosome 1, the AR gene on the X chromosome, and the DLC1, LZTS and MSR1 genes on chromosome 8. The finding of multiple mutations in this latter gene, including clearly inactivating mutations, in both familial and non familial prostate cancer, strongly implicates this gene as an important prostate cancer susceptibility gene. These results implicate for the first time genetic variation affecting macrophage function and the innate immune system as an important determinant of inherited susceptibility for prostate cancer.

References

Carpten J, Nupponen N, Isaacs S, Sood R, Robbins C, Xu J, Faruque M, Moses T, Ewing C, Gillanders E, Hu P, Bujnovszky P, Makalowska I, Baffoe-Bonnie A, Faith D, Smith J, Stephan D, Wiley K, Brownstein M, Gildea D, Kelly B, Jenkins R, Hostetter G, Matikainen M, Schleutker J, Klinger K, Connors T, Xiang Y, Wang Z, Demarzo A, Papadopoulos N, Kallioniemi OP, Burk R, Meyers D, Gronberg H, Meltzer P, Silverman R, Bailey-Wilson J, Walsh P, Isaacs W, Trent J. Germline mutations in the ribonuclease L gene in families showing linkage with HPC1. Nature Genetics 2002; 30:181-184

Chang B, Zheng SL, Isaacs SD, Wiley K, Carpten JD, Hawkins GA, Bleecker ER, Walsh PC, Trent JM, Meyers DA, Isaacs WI, Xu J (2001) Linkage and association of CYP17 gene in hereditary and sporadic prostate cancer. Int J Cancer 95: 354-359

Chang B, Zheng SL, Hawkins GA, Isaacs SD, Wiley KE, Turner A, Carpten JD, Bleecker ER, Walsh PC, Trent JM, Meyers DA, Isaacs WB, Xu J (2002) Joint effect of HSD3B1 and HSD3B2 genes is associated with hereditary and sporadic prostate cancer susceptibility. Cancer Res 62:1784-1789.

Chang B, Zheng SL, Hawkins GA, Isaacs SD, Wiley KE, Turner A, Carpten JD, Bleecker ER, Walsh PC, Trent JM, Meyers DA, Isaacs WB, Xu J (2002) Polymorphic GGC repeats in the androgen receptor gene are associated with hereditary and sporadic prostate cancer risk. Hum Genet 2002; 110:122-129.

Hawkins GA, Mychaleckyj JC, Zheng SL, Faith DA, Kelly B, Isaacs SD, Wiley KE, Chang BL, Ewing CM, Bujnovszky P, Bleecker ER, Walsh PC, Meyers DA, Isaacs WB, Xu J (2002). Germline Sequence Variants of the LZTS1 Gene are Associated with Prostate Cancer Risk. Cancer Genet Cytogenet (In Press).

Ho G, Knapp M, Freije D, Nelson WG, Smith JR, Carpten JD, Bailey-Wilson JE, Beaty TH, Petersen G, Xu J, Kamensky V, Walsh PC, Isaacs WB (2002) Linkage Analyses of Androgen Receptor and Glutathione S-Transferase Pi Variants in Prostate Cancer Families. International Journal of Cancer Int J Cancer 98:938-42.

Smith JR, Freije D, Carpten JD, Grönberg H, Xu J, Issacs SD, Brownstein MJ, Bova GS, et al (1996) Major susceptibility locus for prostate cancer on chromosome 1 suggested by a genome-wide search. Science 274:1371-1374

Xu J, Meyers DA, Freije D, Isaacs S, Wiley K, Nusskern D, Ewing C, et al (1998) Evidence for a prostate cancer susceptibility locus on the X chromosome. Nat Genet 20:175-179

Xu J, Zheng SL, Chang B-L, Smith JR, Carpten JD, Stine OC, Isaacs SD, Wiley KE, Henning L, Ewing CM, Bujnovszky P, Bleecker ER, Walsh PC, Trent JM, Meyers DA, and Isaacs WB (2001) Linkage of prostate cancer susceptibility loci to chromosome 1. Hum Genet 108:335-345

Xu J, Zheng SL, Chang B, Isaacs SD, Wiley K, Hawkin GA, Bleecker ER, Walsh PC, Trent JM, Meyers DA, Isaacs WB (2001) Linkage and association studies of prostate cancer susceptibility gene on 8p22-23. Am J Hum Genet 69:341-350.

Xu J, Zheng SL, Turner A, Isaacs SD, Wiley K, Hawkins GA, Chang B, Bleecker ER, Walsh PC, Meyers DA, Isaacs WB (2002) Associations between hOGG1 sequence variants and prostate cancer susceptibility. Cancer Res 62:2253-2257.

Xu J, Zheng SL, Komiya A, Mychaleckyj JC, Isaacs SD, Hu JJ, Sterling D, Lange EM, Hawkins GA, Turner A, Ewing CM, Faith DA, Johnson JR, Suzuki H, Bujnovszky P, Wiley KE, DeMarzo AM, Bova GS, Chang B, Hall MC, McCullough DL, Partin AW, Kassabian VS, Carpten JD, Bailey-Wilson JE, Trent JM, Ohar J, Bleecker ER, Walsh PC, Isaacs WB, Meyers DA. Germline mutations and sequence variants of the macrophage scavenger receptor 1 gene are associated with prostate cancer risk. Nat Genet 2002 Oct;32(2):321-5

Xu J, Zheng SL, Komiya A, Mychaleckyj JC, Isaacs SD, Chang B, Turner AR, Ewing CM, Wiley KE, Hawkins GA, Bleecker ER, Walsh PC, Meyers DA, Isaacs WB. Common sequence variants of the macrophage scavenger receptor 1 gene are associated with prostate cancer risk. (2003) Am J Hum Genet 72:208-212.

Zheng SL, Mychaleckyj JC, Hawkins GA, Isaacs SD, Wiley KE, Turner A, Chang B, von Kap-Herr C, Carpten JD, Pettenati M, Bleecker ER, Walsh PC, Trent JM, Meyers DA, Isaacs WB, Xu J. Evaluation of DLC1 as a prostate cancer susceptibility gene: mutation screen and association study. Mut Res 2002 (In Press).

APPENDIX

Germline mutations in the ribonuclease L gene in families showing linkage with *HPC1*

J. Carpten[1], N. Nupponen[1], S. Isaacs[2], R. Sood[1], C. Robbins[1], J. Xu[3], M. Faruque[1], T. Moses[1], C. Ewing[2], E. Gillanders[1], P. Hu[1], P. Bujnovszky[2], I. Makalowska[4], A. Baffoe-Bonnie[5,6], D. Faith[2], J. Smith[7], D. Stephan[8], K. Wiley[2], M. Brownstein[9], D. Gildea[1], B. Kelly[2], R. Jenkins[10], G. Hostetter[1], M. Matikainen[11], J. Schleutker[11], K. Klinger[12], T. Connors[12], Y. Xiang[13], Z. Wang[13], A. De Marzo[2], N. Papadopoulos[14], O.-P. Kallioniemi[1], R. Burk[15], D. Meyers[3], H. Grönberg[16], P. Meltzer[1], R. Silverman[13], J. Bailey-Wilson[5], P. Walsh[2], W. Isaacs[2] & J. Trent[1]

Published online: 22 January 2002, DOI: 10.1038/ng823

Although prostate cancer is the most common non-cutaneous malignancy diagnosed in men in the United States[1,2], little is known about inherited factors that influence its genetic predisposition[3–5]. Here we report that germline mutations in the gene encoding 2′-5′-oligoadenylate(2-5A)–dependent RNase L (*RNASEL*)[6–8] segregate in prostate cancer families that show linkage to the HPC1 (hereditary prostate cancer 1) region at 1q24–25 (ref. 9). We identified *RNASEL* by a positional cloning/candidate gene method, and show that a nonsense mutation and a mutation in an initiation codon of *RNASEL* segregate independently in two HPC1-linked families. Inactive *RNASEL* alleles are present at a low frequency in the general population. *RNASEL* regulates cell proliferation and apoptosis through the interferon-regulated 2-5A pathway and has been suggested to be a candidate tumor suppressor gene[10–12]. We found that microdissected tumors with a germline mutation showed loss of heterozygosity and loss of RNase L protein, and that *RNASEL* activity was reduced in lymphoblasts from heterozygous individuals compared with family members who were homozygous with respect to the wildtype allele. Thus, germline mutations in *RNASEL* may be of diagnostic value, and the 2-5A pathway might provide opportunities for developing therapies for those with prostate cancer.

On the basis of reports of familial clustering of prostate cancer[3] and segregation analyses that support the existence of dominantly acting alleles, which confer high risk for prostate cancer[4], we carried out a genetic linkage study using families affected with hereditary prostate cancer (HPC). Results implicated several prostate susceptibility loci, including one on the long arm of chromosome 1, at 1q24–25 (termed HPC1)[9]. We used recombination mapping and candidate gene analysis to map several genes, including *RNASEL*, to the critical region[13,14] of HPC1 (Fig. 1*a*). RNase L is a constitutively expressed latent endoribonuclease that mediates the antiviral and proapoptotic activities of the interferon-inducible 2-5A system[10,11,15]. The gene consists of eight exons. Northern-blot analysis shows that there are two mRNA species of 5 kb and 9.5 kb in the spleen, thymus, prostate, testis, uterus, small intestine, colon and peripheral blood leukocytes (data not shown). Expression varies according to the tissue, with the highest expression in the spleen and thymus.

We initially screened a set of DNA samples representing one affected individual from each of 26 families at high risk for prostate cancer, including 8 families that showed linkage to the HPC1 region and that had at least four affected individuals sharing an HPC1 haplotype. We identified a mutation (Glu265X) in the proband from family 065 (Fig. 1*b*). Four affected brothers had the base substitution 795G→T in exon 2 (starting from the initiating methionine) of *RNASEL*, which is predicted to result in the conversion of a glutamic acid codon to a termination codon at amino-acid position 265 and can lead to the loss of function of that allele[8]. These brothers were heterozygous with respect to the mutation. Three of the four affected brothers had prostate cancers with clinical features that are associated with poor prognosis (that is, Gleason score greater than or equal to 7, stage greater than or equal to T2B, and/or evidence of disseminated disease); this information was not available for the fourth affected brother.

We identified a second mutation in the proband of family 097, a family of African-American descent (Fig. 2*c*). On initial evaluation, five of the six brothers in this family had been diagnosed with prostate cancer; the sixth brother (097-016) was diagnosed subsequently. The average age of diagnosis in this family was 59. The mutation in this family is characterized by the base substitution 3G→A in the codon that corresponds to the initiating methionine (AUG) of the RNase L transcript. This guanine is conserved 100% in the initiation codons of all eukaryotes and in most prokaryotes[16]. This mutation was inherited heterozygously by four of the six affected brothers in family 097. The two affected brothers that do not carry the mutation possibly represent phenocopies (Fig. 2*c*). Whereas three of the four mutation carriers had cancers with poor prognostic indicators, as described above for family 065, the two affected non-mutation carriers had cancers with clinical features that are associated with more favorable disease outcomes (that is, a lower tumor grade and stage, Gleason score 5, and clinically nonpalpable, T1C stage).

[1]*Cancer Genetics Branch, National Human Genome Research Institute, NIH, Bethesda, Maryland 20892, USA.* [2]*Brady Urological Institute, Johns Hopkins Medical Institutions, Baltimore, Maryland, USA.* [3]*Center for Human Genomics, Wake Forest University School of Medicine, Winston-Salem, North Carolina, USA.* [4]*Genome Technology Branch, and* [5]*Genetic Disease Research Branch, National Human Genome Research Institute, NIH, Bethesda, Maryland, USA.* [6]*Division of Population Sciences, Fox Chase Cancer Center, Philadelphia, Pennsylvania, USA.* [7]*Division of Genetic Medicine, Vanderbilt University, Nashville, Tennessee, USA.* [8]*Research Center for Genetic Medicine, Children's National Medical Center, Washington, DC, USA.* [9]*Laboratory of Genetics, National Institute of Mental Health, NIH, Bethesda, Maryland, USA.* [10]*Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, Minnesota, USA.* [11]*Laboratory of Cancer Genetics, Institute of Medical Technology, University of Tampere and Tampere University Hospital, Tampere, Finland.* [12]*Genzyme Molecular Oncology, Framingham, Massachusetts, USA.* [13]*Department of Cancer Biology, Lerner Research Institute, The Cleveland Clinic Foundation, Cleveland, Ohio, USA.* [14]*Institute of Cancer Genetics, Department of Pathology, Columbia University, New York, New York, USA.* [15]*Department of Microbiology and Immunology, Albert Einstein School of Medicine, Yeshiva University, Bronx, New York, USA.* [16]*Department of Oncology, Umeå University, Umeå, Sweden. Correspondence should be addressed to J.T. (e-mail: jtrent@nih.gov).*

To assess the frequency of the Glu265X and the Met1Ile mutations, we analyzed the DNA from control populations and from individuals with non-familial prostate cancer (Table 1). We found one Glu265X heterozygote in 144 normal control individuals and two Glu265X heterozygotes in 186 participants with no family history of prostate cancer and normal serum concentrations of prostate-specific antigen. Analysis of germline DNA from 258 men with non-familial prostate cancer revealed two Glu265X heterozygotes. Thus, this nonsense variant is found in the control population at an estimated allele frequency of 0.5%; as yet we cannot identify a difference in allele frequency between affected individuals and controls.

By contrast, we did not observe the Met1Ile mutation in 698 control individuals, 284 of whom were African American (Table 1). We note that in rare, highly penetrant disorders with a young age of onset, it is unusual to find mutant alleles in unaffected individuals. For low penetrance disorders with a late age of onset and extremely high prevalence (for example, prostate cancer), however, it is common to identify individuals in a general control population who may be carriers of a mutant allele[17,18]. In addition to these two mutations, we also identified a series of missense mutations in the probands of HPC-affected individuals (Web Fig. A). Studies are underway to determine the frequency of these changes in affected individuals and controls.

Single-strand conformation polymorphism (SSCP) analysis of microdissected tumor DNA from individual 065-009 showed that, compared with the heterozygosity of their normal lymphocyte DNA, there was clear loss of the wildtype allele in tumor cells from this individual (Fig. 2*a*). We used the same tumor sample from individual 065-009 in an immunohistochemical analysis of RNase L protein expression using a monospecific RNase L monoclonal antibody[19]. Although we observed cytoplasmic staining in non-cancerous prostate epithelial cells, there was a consistent absence of staining in cancer cells throughout the tumor, consistent with inactivation of both *RNASEL* alleles in tumor tissue from this individual (Fig. 2*b*).

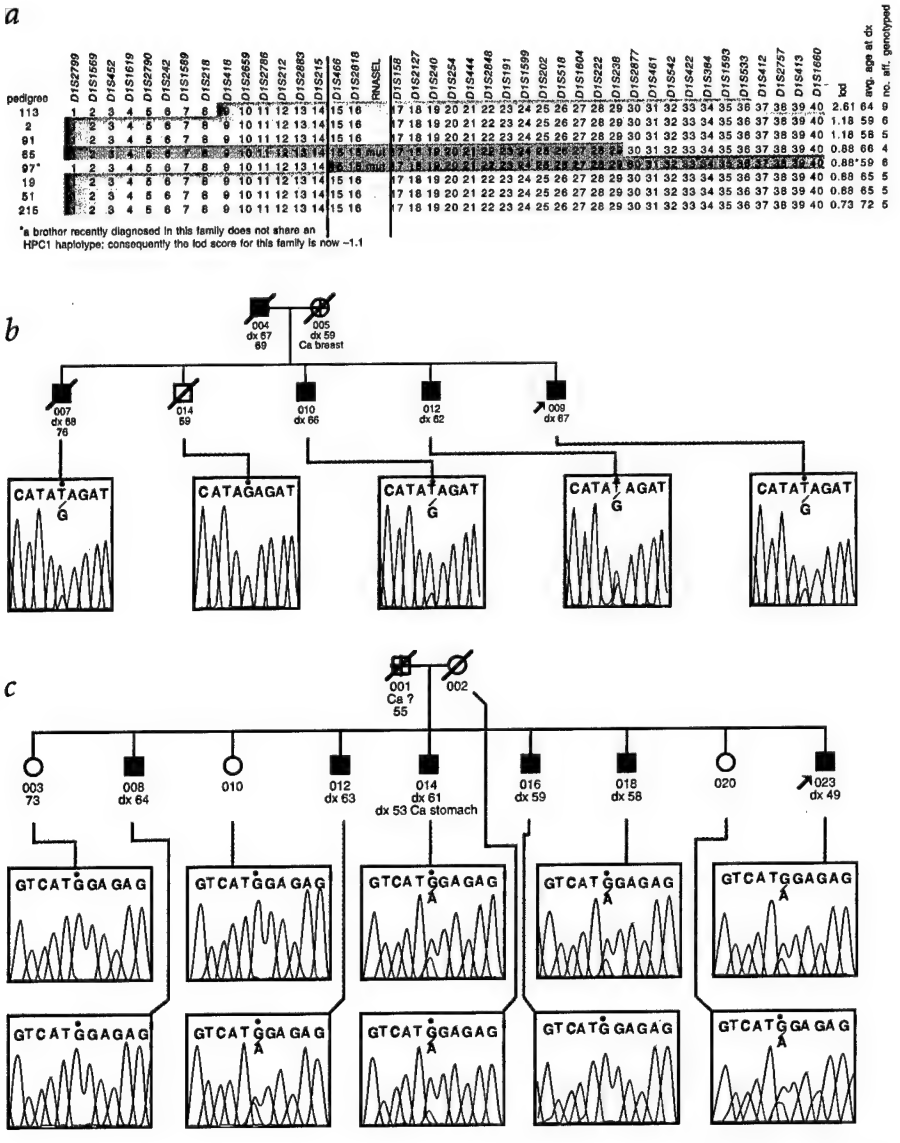


Fig. 1 Multipoint linkage analysis assuming heterogeneity on 91 high-risk prostate cancer families suggests *HPC1* maps to *D1S2883–D1S158–D1S422*. ***a***, Informative affected recombinants, prioritized on the basis of the individual lod scores of families, led to the identification of a critical interval for *HPC1*. ***b,c***, Detailed family structure and sequence chromatograms showing transmission of the Glu265X mutation in family 065 and the Met1Ile mutation in family 097. The variable nucleotide is marked by a dot in each representative chromatogram.

To determine whether the Glu265X mutation affected enzyme activity, we measured rRNA cleavage in lymphoblasts after transfections with the biostable 2-5A analog psA(2′ps5′A)$_3$ (Fig. 2*c* and Table 2). The prostate cancer cell line PC3, which originated from a bone metastasis of a grade IV prostatic adenocarcinoma in a 62-year-old male who lacked the described mutations in *RNASEL*, showed prominent products specific to RNase L cleavage[20] of 28S and 18S rRNA (Fig. 2*c*, lanes 1–3). The lymphoblasts had less activity, partly owing to lower uptake of 2-5A, as determined by fluorescein-tagged 2-5A and confocal microscopy (data not shown). Lymphoblasts from heterozygous individuals from family 065, including 065-012, showed decreased *RNASEL* activity compared with lymphoblasts from individuals homozygous with respect to the wildtype allele, for example 065-016 (Fig. 2*c*, lanes 4–9, and Table 2). Three separate experiments on the lymphoblasts from family 065 resulted in homozygous/heterozygous ratios of *RNASEL* activity towards 18S and 28S rRNAs of 2.2 and 1.9, respectively (Table 2). Similarly, several experiments with lymphoblasts from family 097 produced homozygous/heterozygous ratios of *RNASEL* activity towards 18S and 28S rRNA of 2.4 and 1.5, respectively (Table 2). The average of the ratios from both families (2.0) suggests that, as predicted, homozygous cells contain twice the amount of RNase L as do heterozygous cells. These findings also indicate that both the Glu265X and Met1Ile mutations prevent synthesis of a functional RNase L.

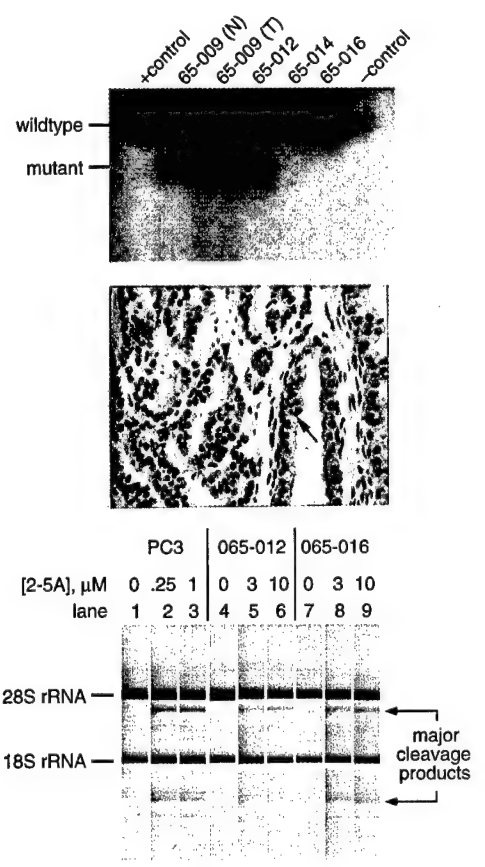


Fig. 2 Deficiences in RNase L in tumor tissue and lymphoblasts from HPC1-affected individuals. ***a***, Loss of heterozygosity using SSCP analysis was carried out on lymphocyte DNA from individual 065-009 and controls, and on tumor DNA from individual 065-009 to determine loss of the wildtype allele in tumors. Wildtype and mutant alleles are indicated. ***b***, Immunohistochemical analysis of the expression of RNAse L protein[19] in a tumor specimen from an Glu265X mutation carrier in family 065. The cytoplasm of normal prostate epithelium stains positively (brownish red stain in cells marked by the arrow on the right of the section), whereas tumor cells are negative (arrowhead). ***c***, Activity of *RNASEL* in intact PC3 cells and lymphoblast cell lines from individual 065-012 (heterozygous with respect to the Glu265X mutation in *RNASEL*) and individual 065-016 (homozygous, lacking this mutation). The positions of the 28S and 18S rRNA and their main cleavage products are indicated.

It has been proposed that *RNASEL* is a candidate tumor-suppressor gene on the basis of its known function[12]. RNase L has been shown to be lost completely in the hepatoma cell line HepG2 (ref. 21). An animal model of RNase L function shows that mice devoid of RNase L have defects in both interferon-induced apoptosis and antiviral response[11]; however, these animals do not develop tumors. Although the function of the 2-5A pathway has not yet been explored in prostate tissue, the balance between hormonally regulated growth and cell death is crucial in this organ[22]. It is possible that reduced 2-5A function shifts this balance toward cell growth, creating a favorable environment for the development of prostate cancer.

We propose that *RNASEL* is a candidate cancer-susceptibility gene for HPC1. We have identified two mutations that are potentially responsible for prostate cancer cases in families showing linkage to the HPC1 locus. We have also shown that loss of the wildtype allele occurs in tumor DNA from a mutation carrier, suggesting that there is complete loss of function of this protein. The low frequency of mutations suggests that there is likely to be a high rate of heterogeneity in prostate cancer. It is also possible that we did not identify mutations in other potential families with HPC1 that we screened owing to either the location of the mutations in regulatory elements or technical limitations. The identification of other functionally significant mutations in *RNASEL* using an independent data set will be necessary to confirm this gene as the prostate cancer–susceptibility gene in families showing linkage to HPC1. These findings could be significant, as the identification of germline mutations in this gene could lead to early diagnosis and therapeutic approaches for prostate cancer cases linked to HPC1.

Methods

Subjects. We obtained informed consent from each participant in this study.

PCR analysis. The primers for PCR are available upon request. We carried out PCR of *RNASEL* exons in a volume of 50 μl containing 20 ng of genomic DNA, PCR buffer (Gibco BRL), 2.25 mM Mg^{2+}, 250 nM dNTPs, 10 pmol of each forward/reverse primer mix, 0.06 U Platinum Taq DNA polymerase (Gibco BRL) and 0.06 U AmpliTaq Gold (PE Biosystems).

The PCR protocol was 95 °C for 14 min, 35 cycles of 95 °C for 30 s, 56 °C for 30 s, and 72 °C for 45 s, and a final extension at 72 °C for 10 min. We analyzed PCR products by 2% agarose gel electrophoresis.

DNA sequencing and sequence analysis. We amplified the exons for a given gene using M13-tailed primers. The subsequent PCR products were cleaned up using a PCR purification kit (Qiagen) and the BIOROBOT 9600 dual vacuum system (Qiagen). We prepared half-volume cycle sequencing reactions in 96-well plates using standard M13 forward and reverse primers and 3700 Big Dye Terminator Chemistry (PE/Applied Biosystems). After purification, sequencing reactions were run on a 3700

Table 1 • Frequency of Glu265X and Met1Ile mutations in familial prostate cancer cases, nonfamilial cases and controls

Mutation	Study groups	Number screened	Number of mutants	Frequency (%)	Sample type	Method
Glu265→X						
	unaffected white men*	186	2	0.54	blood DNA	direct sequencing and SSCP
	CEPH parents	96	1	0.52	blood DNA	direct sequencing
	US population controls	48	0	0.00	blood DNA	
		330	3	0.45		
	prostate cancer cases (non-HPC)	258	2	0.39	normal tissue DNA	SSCP
Met1→Ile	unaffected African American men*	92	0	0.00	blood DNA	*Nla*III digest
	unaffected white men*	186	0	0.00	blood DNA	direct sequencing
	African American population control	192	0	0.00	blood DNA	direct sequencing
	US population controls	48	0	0.00	blood DNA	direct sequencing
	total number of samples screened	240	0	0.00		
	prostate cancer cases (non-HPC)	180	0	0.00	normal tissue DNA	*Nla*III digest

*From prostate cancer screenings (PSA<4.0, ages 35–70, men were excluded if they had an abnormal digital rectal exam). CEPH, Centre d'Etude du Polymorphisme Humain.

Table 2 • RNase L activity in intact lymphoblasts

Cell line	18S rRNA cleavage (%)	28S rRNA cleavage (%)
RNase $L^{+/+}$		
065-016 (*n*=3)	15.3 ± 1.5	11.2 ± 0.8
065-061 (*n*=3)	12.3 ± 2.7	11.3 ± 2.2
Average	13.8	11.3
RNase $L^{+/-}$		
065-007 (*n*=3)	5.7 ± 4	6.8 ± 3.6
065-009 (*n*=3)	8.2 ± 2	6.1 ± 1.5
065-012 (*n*=4)	5.0 ± 2.1	4.8 ± 1.7
Average	6.3	5.9
RNase $L^{+/+}$/ RNase $L^{+/-}$	2.2-fold	1.9-fold
RNase $L^{+/+}$		
097-008 (*n*=6)	33.1 ± 4.4	15.7 ± 3.1
097-016 (*n*=6)	26.7 ± 5.6	14.2 ± 2.1
Average	29.9	15.0
RNase $L^{+/-}$		
097-012 (*n*=3)	18.5 ± 6.6	15.3 ± 7.1
097-014 (*n*=5)	7.7 ± 1.6	9.8 ± 5.1
097-018 (*n*=5)	11.5 ± 3.9	8.3 ± 3.2
097-023 (*n*=5)	12.0 ± 4.3	6.0 ± 2.2
Average	12.4	9.9
RNase $L^{+/+}$/ RNase $L^{+/-}$	2.4-fold	1.5-fold

DNA Analyzer (PE/Applied Biosystems) according to the manufacturer's protocols. We aligned and analyzed sequence chromatograms using Sequencher version 4.1 (Gene Codes).

Laser capture microdissection. Glass slides containing sections 8 mm thick cut from paraffin-embedded tissue blocks were deparaffinized and stained with hematoxylin and eosin. We dehydrated slides in xylene and used them immediately for laser capture microdissection of tumor cells using the PixCell II LCM system (Arcturus). A polymer cap was placed on the slide, and all available tumor cells (~2,000) were transferred to the cap using a laser beam. We then placed the cap on an Eppendorf tube containing 50 ml of digestion buffer (1 mg ml^{-1} proteinase K, 10 mM Tris-HCl, pH 8, 1 mM EDTA and 1% Tween-20). We incubated the solution at 52 °C overnight with the tube in an inverted position so that digestion buffer was in contact with the tissue on the cap. The cap was removed after centrifugation for 5 min, and proteinase K was inactivated by incubation at 95 °C for 10 min.

Loss of heterozygosity. We determined loss of heterozygosity using SSCP analysis for the Glu265X mutation. An expected 166-bp product spanning the Glu265X mutation was identified by PCR. We modified the PCR protocol such that the volume was reduced to 15 ml, [a-^{32}P]dCTP was added, and 40 cycles were carried out. The PCR products were mixed with formamide loading buffer, denatured and separated by electrophoresis on a Hydrolink MDE gel (BioWhittaker Molecular Applications) with 5% glycerol for 16 h. We dried the gels in a vacuum and subjected them to autoradiography.

***RNASEL* activity assay.** Lymphoblastoid cell lines were cultured in RPMI 1640 medium supplemented with glutamine and 15% fetal bovine serum. The PC3 prostate cancer cell line[23] was grown in the same medium, except that 10% fetal bovine serum was used. The biostable, all phosphorothioate mixed isomer analog of tetramer 2-5A, psA(2′ps5′A)$_3$ was synthesized chemically using an ABI 380B DNA synthesizer, purified by high-performance liquid chromatography and desalted (Z. Wang and R.H.S., unpublished data). We tranfected cells with 3.0 μM psA(2′ps5′A)$_3$ or at the indicated concentrations for 4–5 h using lipofectamine (Gibco-BRL). We isolated total RNA from transfected cells using Trizol reagent (Gibco-BRL) and quantified it by measuring absorbance at 260 nm. We separated RNA molecules on RNA chips and analyzed them with an Agilent Bioanalyzer 2100 (Agilent Technologies). We determined peak areas of 28S and 18S rRNA and their main cleavage products using the Bio Sizing (version A.01.30 S1220) program (Agilent Technologies).

Note: Supplementary information is available on the Nature Genetics web site (http://genetics.nature.com/supplementary_info/).

Acknowledgments

We wish to thank the affected individuals and their family members who made this study possible. We thank D. Freije, H. Suzuki, E. Wilkens, A. Kibel, G. Bova, S. Gregory and T. Bonner for contributions to earlier phases of this work; F. S. Collins for input and comments; J. Qian for FISH analyses; M. Emmert-Buck for help in laser-capture microdissection; J. Hicks for immunohistochemistry; and J.R. Okicki for synthesizing the fluorescein-tagged 2-5A. This work was supported in part by grants from the PHS (SPORE), DOD, CaPCURE (W.I.) and the Fund for Research and Progress in Urology, the Johns Hopkins University, the Swedish Cancer Society and the SSF Genome Program (H.G.), the V Foundation for Cancer Research (J.S.), the Finnish Cultural Foundation, the Helsingin Sanomat Foundation, the Paulo Foundation, the Ella & Georg Ehrnrooth Foundation and the Maud Kuistila Foundation (N.N.), the NIH (R.J. and R.H.S.), and DOD (J.X.).

Competing interests statement. *The authors declare that they have no competing financial interests.*

Received 5 November; accepted 16 November 2001.

1. Hankey, B.F. *et al.* Cancer surveillance series: interpreting trends in prostate cancer—part I: evidence of the effects of screening in recent prostate cancer incidence, mortality, and survival rates. *J. Natl Cancer Inst.* **91**, 1017–1024 (1999).
2. Dennis, L.K. & Resnick, M.I. Analysis of recent trends in prostate cancer incidence and mortality. *Prostate* **42**, 247–252 (2000).
3. Steinberg, G.D., Carter, B.S., Beaty, T.H., Childs, B. & Walsh, P.C. Family history and the risk of prostate cancer. *Prostate* **17**, 337–347 (1990).
4. Carter, B.S., Beaty, T.H., Steinberg, G.D., Childs, B. & Walsh, P.C. Mendelian inheritance of familial prostate cancer. *Proc. Natl Acad. Sci. USA* **89**, 3367–3371 (1992).
5. Ostrander, E.A. & Stanford, J.L. Genetics of prostate cancer: too many loci, too few genes. *Am. J. Hum. Genet.* **67**, 1367–1375 (2000).
6. Clemens, M.J. & Williams, B.R. Inhibition of cell-free protein synthesis by pppA2′p5′A2′p5′A: a novel oligonucleotide synthesized by interferon-treated L cell extracts. *Cell* **13**, 565–572 (1978).
7. Floyd-Smith, G., Slattery, E. & Lengyel, P. Interferon action: RNA cleavage pattern of a (2′-5′)oligoadenylate-dependent endonuclease. *Science* **212**, 1030–1032 (1981).
8. Zhou, A., Hassel, B.A. & Silverman, R.H. Expression cloning of 2-5A-dependent RNAase: a uniquely regulated mediator of interferon action. *Cell* **72**, 753–765 (1993).
9. Smith, J.R. *et al.* Major susceptibility locus for prostate cancer on chromosome 1 suggested by a genome-wide search. *Science* **274**, 1371–1374 (1996).
10. Hassel, B.A., Zhou, A., Sotomayor, C., Maran, A. & Silverman, R.H. A dominant negative mutant of 2-5A-dependent RNase suppresses antiproliferative and antiviral effects of interferon. *EMBO J.* **12**, 3297–3304 (1993).
11. Zhou A. *et al.* Interferon action and apoptosis are defective in mice devoid of 2′,5′- oligoadenylate-dependent RNase L. *EMBO J.* **16**, 6355–6363 (1997).
12. Lengyel, P. Tumor-suppressor genes: news about the interferon connection. *Proc. Natl Acad. Sci. USA* **90**, 5893–5895 (1993).
13. Carpten, J.D. *et al.* A 6-Mb high-resolution physical and transcription map encompassing the hereditary prostate cancer 1 (HPC1) region. *Genomics* **64**, 1–14 (2000).
14. Sood, R. *et al.* Cloning and characterization of 13 novel transcripts and the human rgs8 gene from the 1q25 region encompassing the hereditary prostate cancer (hpc1) locus. *Genomics* **73**, 211–222 (2001).
15. Kerr, I.M. & Brown, R.E. pppA2′p5′A2′p5′A: an inhibitor of protein synthesis synthesized with an enzyme fraction from interferon-treated cells. *Proc. Natl Acad. Sci. USA* **75**, 256–260 (1978).
16. Sherman, F., McKnight, G. & Stewart, J.W. AUG is the only initiation codon in eukaryotes. *Biochim. Biophys. Acta.* **609**, 343–346 (1980).
17. Hartge, P., Struewing, J.P., Wacholder, S., Brody, L.C. & Tucker, M.A. The prevalence of common BRCA1 and BRCA2 mutations among Ashkenazi Jews. *Am. J. Hum. Genet.* **64**, 963–970 (1999).
18. Ogura, Y. *et al.* A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. *Nature* **411**, 603–606 (2001).
19. Dong, B. & Silverman, R.H. 2-5A-dependent RNase molecules dimerize during activation by 2-5A. *J. Biol. Chem.* **270**, 4133–4137 (1995).
20. Silverman, R.H., Skehel, J.J., James, T.C., Wreschner, D.H. & Kerr, I.M. rRNA cleavage as an index of ppp(A2′p)nA activity in interferon-treated encephalomyocarditis virus-infected cells. *J. Virol.* **46**, 1051–1055 (1983).
21. Tnani, M. & Bayard, B.A. Lack of 2′,5′-oligoadenylate-dependent RNase expression in the human hepatoma cell line HepG2. *Biochim. Biophys. Acta* **1402**, 139–150 (1998).
22. Isaacs, J.T., Furuya, Y. & Berges, R. The role of androgen in the regulation of programmed cell death/apoptosis in normal and malignant prostatic tissue. *Semin. Cancer Biol.* **5**, 391–400 (1994).
23. Kaighn, M.E., Narayan, K.S., Ohnuki, Y., Lechner, J.F. & Jones, L.W. Establishment and characterization of a human prostatic carcinoma cell line (PC-3). *Invest. Urol.* **17**, 16–23 (1979).

Int. J. Cancer (Pred. Oncol.): 95, 354–359 (2001)

Publication of the International Union Against Cancer

LINKAGE AND ASSOCIATION OF CYP17 GENE IN HEREDITARY AND SPORADIC PROSTATE CANCER

Bao-li CHANG[1], Siqun L. ZHENG[2], Sarah D. ISAACS[3], Kathy E. WILEY[3], John D. CARPTEN[4], Gregory A. HAWKINS[2], Eugene R. BLEECKER[2], Patrick C. WALSH[3], Jeffrey M. TRENT[4], Deborah A. MEYERS[2], William B. ISAACS[3] and Jianfeng XU[2]*

[1]*University of Maryland School of Medicine, Baltimore, MD, USA*
[2]*Center for Human Genomics, Wake Forest University School of Medicine, Winston-Salem, NC, USA*
[3]*Department of Urology, Johns Hopkins Medical Institutions, Baltimore, MD, USA*
[4]*National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA*

Androgens are essential for prostate development, growth and maintenance and the association between androgen levels and prostate cancer is well established. Since the CYP17 gene encodes the enzyme cytochrome P450c17α, which mediates 17α-hydroxylase and 17,20-lyase activities in the androgen biosynthesis pathway, sequence variations in the gene and association with increased risk to prostate cancer has been studied. In particular, several groups have studied the association between a polymorphism in the 5′ promoter region and prostate cancer using a population-based association approach. However, the results from these studies were inconclusive. To further study this polymorphism and its possible role in hereditary prostate cancer (HPC), we performed a genetic linkage analysis and family-based association analysis in 159 families, each of which contains at least 3 first-degree relatives with prostate cancer. In addition, we performed a population-based association analysis to compare the risk of this polymorphism to hereditary and sporadic prostate cancer in 159 HPC probands, 249 sporadic prostate cancer patients and 211 unaffected control subjects. Evidence for linkage at the CYP17 gene region was found in the total 159 HPC families (LOD = 1.3, $p = 0.01$, at marker D10S222). However, family-based association tests did not provide evidence for overtransmission of either allele of the CYP17 polymorphism to affected individuals in the HPC families. The allele and genotype frequencies of the polymorphism were not statistically different among the HPC probands, sporadic cases and unaffected control subjects. In conclusion, our results suggest that the CYP17 gene or other genes in the region may increase the susceptibility to prostate cancer in men; however, the polymorphism in the 5′ promoter region has a minor role if any in increasing prostate cancer susceptibility in our study sample.

Key words: *prostate cancer; linkage; association; genetic; hereditary; CYP17*

Prostate cancer (MIM 176807) is the most frequently diagnosed noncutaneous cancer in men in Western countries. Both genetic and environmental factors may be involved in the etiology of prostate cancer. The well-established risk factors for prostate cancer include age, race-ethnicity and family history. Men with 2 or more affected first-degree relatives are 5 to 11 times more likely to develop prostate cancer.[1] The genetic component of the familial clustering of prostate cancer has been demonstrated in several twin and segregation studies.[2–7] So far, several prostate cancer susceptibility loci, including HPC1 (MIM 601518) at 1q24–25, PCAP (MIM 602759) at 1q42–43, HPCX (MIM 300147) at Xq27-28, CAPB (MIM 603688) at 1p36 and HPC20 at 20q13, have been mapped based on fine mapping linkage studies.[8–13] It is believed that about 9% of all prostate cancer cases are due to mutations in prostate cancer susceptibility genes.[2]

Androgens, a group of steroid hormones that are essential for prostate development, growth and maintenance, have been hypothesized to be involved in prostate carcinogenesis. Androgen ablation remains the most effective therapy for the treatment of advanced prostate cancer.[14–16] A strong trend toward increasing prostate cancer risk with increasing levels of plasma testosterone has been observed.[17] In addition, African-Americans, a population at high risk for prostate cancer, were found to have higher levels of plasma testosterone than Caucasian Americans, a population at lower risk for prostate cancer.[18] The same hormone-promoting carcinogenesis phenomenon was also observed in other types of cancers.

It is believed that the effects of androgens as well as the inherited genotypes that regulate androgen metabolism may modify an individual's risk of prostate cancer. Several genes in the androgen metabolism pathway, including the cytochrome P450c17α (CYP17) gene, have been proposed and tested as candidate genes for prostate cancer. The enzyme P450c17α mediates both 17α-hydroxylase and 17,20-lyase activity. It catalyzes both 17α-hydroxylation of pregnenolone and progesterone and 17,20-lysis of 17α-hydroxypregnenolone and 17α-hydroxyprogesterone, which are the key reactions for both sex steroid and cortisol biosynthesis. The CYP17 gene was mapped to 10q24.3 and consists of 8 exons.[19] Mutations in the CYP17 gene result in disrupted testosterone synthesis that leads to pseudohermaphroditism in the male and impaired sex steroid hormone synthesis and absence of sexual maturation in the female. A single-base polymorphism (a T to C transition) in the 5′ untranslated region of the CYP17 gene was identified.[20] The single-base change creates an additional SP1-type (CCACC box) transcriptional factor binding site that was hypothesized to result in increased expression of the gene. This polymorphism also creates a recognition site for the restriction enzyme MspA1. Following MspA1 digestion of a PCR fragment, the A1 allele (representing the wild-type allele) and A2 allele (representing the variant allele with C transition) were designated. Many population-based association studies have been conducted to investigate the possible effects of this polymorphism on the risk of hormone-related cancers. The A2 allele has been associated with elevated serum progesterone and estrogen levels in women[21,22] and is related to an increased risk of advanced breast cancer,[23] early-onset breast cancer[24] and male breast cancer.[25] However, other studies failed to demonstrate an association between the A2 allele and the risk of breast cancer or steroid hormone levels.[26–29] Contradictory results regarding the CYP17 genotype and increased risk for prostate cancer were also reported. Lunn *et al.*,[30] Gsur *et al.*[31] and Yamada *et al.*[32] reported positive associations between the A2 allele and an increased risk for prostate cancer. However, 2 other studies[33,34] showed that the A1 allele is the risk allele for prostate cancer. The association between the risk for prostate cancer and the CYP17 genotype remains controversial even in age-stratified

Grant sponsor: PHS SPORE; Grant number: CA58236; Grant sponsor: US Department of Defense.

*Correspondence to: Center for Human Genomics, Wake Forest University School of Medicine, Medical Center Blvd., Winston-Salem, NC 27157, USA. Fax: +336-716-7575. E-mail: jxu@wfubmc.edu

Received 21 March 2001; Revised 6 June 2001; Accepted 18 June 2001

subgroups. Two groups observed an association between the CYP17 genotype and prostate cancer risk in younger age of onset prostate cancer cases.[30,32] On the other hand, the association between the CYP17 genotype and prostate cancer in the older age of onset subgroup was documented in 2 other publications.[31,34]

The ability to detect a relationship between a mutation (polymorphism) and disease susceptibility is dependent on its frequency and penetrance. While a population-based association study is the method of choice to detect mutations with a high frequency but low penetrance, genetic linkage studies and family-based association studies in families with multiple affected members are better study designs for detecting mutations with a low frequency but high penetrance. This is because the frequency of the gene carriers is likely to be higher in these families, and there is a higher likelihood that a rare mutation cosegregates with disease in family members carrying the disease. Because the underlying influence of the mutation in the CYP17 gene is unknown, *a priori,* approaches utilizing multiple study designs are needed to evaluate the gene in the etiology of prostate cancer.

In our study, we conducted a genetic linkage study, family-based association study and population-based association study to evaluate the CYP17 gene in prostate cancer. The linkage study was performed in 159 HPC families, which allowed us to test whether the genes in the region of CYP17 increase prostate cancer susceptibility. The family-based association study was conducted in the same 159 HPC families, which permitted us to test whether the 5′ promoter polymorphism in the CYP17 gene is associated with prostate cancer. A population-based association study was performed in the 159 HPC probands, 249 sporadic prostate cancer cases and 211 unaffected controls, which allowed us to test whether the polymorphism increases the risk for sporadic and hereditary prostate cancer.

MATERIAL AND METHODS

Subjects

A detailed description of the study sample was presented elsewhere.[35] Briefly, a total of 159 HPC families were collected and studied at the Brady Urology Institute at Johns Hopkins Hospital (Baltimore, MD). The diagnosis of prostate cancer was verified by medical records for each affected male studied. Age of diagnosis of prostate cancer was confirmed either through medical records or from 2 other independent sources. The mean age at diagnosis was 64.3 years. Eighty-four percent of the families were Caucasian, 6.9% were Ashkenazi Jewish and 8.8% were African-American. The average number of affected men per family was 5.08. The numbers of families with 3, 4 and ≥5 affected men were 29, 40 and 90, respectively.

All 249 unrelated prostate cancer cases were recruited from patients who underwent treatment for prostate cancer at the Johns Hopkins Hospital. The diagnosis of prostate cancer for all these subjects was confirmed by pathology reports. Preoperative prostate-specific antigen (PSA) levels, Gleason score and pathologic stages were available for 92, 244 and 245 of the 249 cases, respectively. Mean age at diagnosis for these cases was 58.6 years, but family history information was not obtained. More than 93% of the cases are Caucasian and 3.2% are African-American.

Two hundred twenty-two nonprostate cancer controls were selected from men participating in screening programs for prostate cancer. By applying the exclusion criteria of abnormal digital rectal examination (DRE) and abnormal PSA level (*i.e.,* ≥4 ng/ml), 211 were eligible for the study. The mean age at examination was 58 years. More than 86% of the eligible controls are Caucasian and 7.1% are African-American. About 5.6% of the eligible controls have brothers or a father affected with prostate cancer (based on interview of the controls).

All individuals in our study gave full, informed consent.

Genotyping methods

Two microsatellite markers, D10S192 and D10S222, surrounding the CYP17 gene were genotyped in 159 HPC families. These markers were selected from Marshfield Comprehensive Human Genetic Maps,[36] which span approximately 1 cM surrounding CYP17 gene at chromosomal region 10q24.3. The order and distances are D10S192 at 102.02 cM, CYP17 at 102.65 cM and D10S222 at 103.03 cM, based on the LDB map.[37] Multiplex PCR using fluorescently labeled primers (fam or hex) was performed, and the resulting PCR fragments were separated on an ABI 3700 sequencer. The genotypes were scored using ABI software (Genotyper). A modified version of the program Linkage Designer (http://dnalab-www.uia.ac.be/dnalab/ld.html) was used to bin the alleles and check inheritance. The output from Linkage Designer was then analyzed further for any inconsistencies by running the LINKAGE software[38,39] without disease phenotype information. Marker allele frequencies were estimated from the unrelated individuals (pedigree founders) from the 159 HPC pedigrees for which genotype information was available.

The single nucleotide polymorphism (SNP) of the C to T transition in CYP17 was detected using polymerase chain reaction (PCR) and restriction fragment length polymorphism (RFLP) analysis. PCR amplification of the 209 bp DNA fragment in the 5′ region of the CYP17 gene was performed using the primers CYP17F (5′-GGC TCC AGG AGA ATC TTT C-3′) and CYP17R (5′-GGG CCA AAA CAA ATA AGC TA-3′). PCR reactions were carried out in 10 μl aliquots containing 30 ng of genomic DNA, 0.5 μM of each primer, 50 mM KCl, 10 mM Tris, 100 uM dNTPs, 1.5 mM MgCl and 0.5 unit of Taq polymerase (GIBCO BRL). The PCR reaction consisted of an initial 4 min denaturation step followed by 33 cycles of 30 sec at 94°C, 30 sec at 60°C and 20 sec at 72°C. The PCR products were incubated with the restriction enzyme Msp AI (New England Biolabs) for 2 hr at 37°C and run on 2% agarose gels. MspAI cuts in the presence of an additional Sp-1-type promoter site (A2 allele). The digested products were 123 and 86 bp.

Statistical methods

Hardy-Weinberg Equilibrium (HWE) tests were performed using the software package GDA (http://lewis.eeb.uconn.edu/lewishome/gda.html).[40] A large number (10,000) of the possible arrays were generated by permuting the alleles among genotypes, and the proportion of these permuted genotypic arrays that have a smaller conditional probability than the original data were calculated (empirical *p*-values).

Multipoint linkage analyses were performed using both parametric and nonparametric methods, implemented by the computer program GENEHUNTER-PLUS.[41,42] For the parametric analysis, the same autosomal dominant model that was used by Smith *et al.*[8] was assumed. Linkage in the presence of heterogeneity was assessed by use of Smith's admixture test for heterogeneity.[43] A maximum likelihood approach was used to estimate the proportion of linked families (α) by maximizing the admixed LOD score (HLOD).

For the nonparametric analysis, the estimated identical by descent (IBD) sharing of alleles for the various affected relative pairs was compared to its expected values under the null hypothesis of no linkage. A statistic Z-all in the program was used.[44] Allele sharing LOD scores were then calculated based on the statistic Z-all and assigning equal weight to all families using the computer program ASM.[42]

Both HLOD and allele sharing LOD scores can be converted to a χ^2 ($\chi^2 = 4.6 \times$ LOD score). Although the true distribution of the χ^2 under the null hypothesis of no linkage is unknown, especially in the situation of multipoint analysis, we assume that the distribution is a mixture of one that is degenerate at zero and one that can be approximated by the distribution of the maximum of 2 independent χ^2 variables, each with 1 degree of freedom.[45] *p*-

values were thus calculated by $0.5*(1-(1-p_1)(1-p_1))$, where p_1 is the p-value of a χ^2 with 1 degree of freedom.

Family-based association tests were performed for the polymorphism and microsatellite markers in the 159 HPC families, using a software package FBAT.[46] Unlike the classic transmission disequilibrium test (TDT), which is limited to a specific pedigree structure (1 genotyped proband and 2 genotyped parents per pedigree), the FBAT utilizes data from nuclear families, sibships, or a combination of the 2, to test for linkage and linkage disequilibrium (association) between traits and genotypes. The test for linkage is valid when multiple affected members per pedigree are used, and the power to detect linkage is increased if there is an association. The test for association is valid if 1 affected member per pedigree is used (the genotypes of all the affected members can be included), or if the empirical variance is used to account for correlation between transmissions in families when linkage is present. Briefly, the FBAT determines an S statistic from the data, which is the linear combination of offspring genotypes and phenotypes. The distribution of the S statistics is generated by treating the offspring genotype data as random and conditioning on the phenotypes and parental genotypes. When the marker is biallelic, a Z statistic and its corresponding p-value are calculated. When the marker is multiallelic, a χ^2 test is performed, with number of degrees of freedom equal to the number of alleles.

Population-based association tests were performed for the polymorphism in prostate cancer cases and unaffected controls. An unconditional logistic regression is used to test for association between genotypes and affection status, adjusting for potential confounders such as age.

RESULTS

Linkage study at CYP17 gene region in 159 HPC families

The 2 microsatellite markers (D10S192 and D10S222) and the 5′ promoter polymorphism in the CYP17 gene were genotyped in the available DNA samples from the 159 HPC families. The markers and the SNP were in HWE in 159 probands. Pair-wise linkage disequilibrium tests were not significant, suggesting that they are in linkage equilibrium. The heterozygosity scores for D10S192, CYP17 and D10S222 were 0.83, 0.37 and 0.65, respectively. The multipoint parametric analysis provided evidence for linkage with a peak HLOD of 1.30 ($p = 0.014$) at D10S222. There were 33 families with LOD scores ≥0.3. The nonparametric analysis provided an NPL score of 1.04 ($p = 0.14$) and an allele sharing LOD of 0.46 ($p = 0.13$) at the marker D10S222 (Table I). Stratified linkage analyses were also performed based on family characteristics such as mean age of diagnosis, number of affected members in the family and ethnicity (Table 1). The evidence for linkage at the CYP17 gene region is mainly from the families with mean age of diagnosis ≥65 years ($n = 80$, HLOD = 1.21, $p = 0.018$), the families with ≥5 affected members ($n = 90$, HLOD = 1.13, $p = 0.022$) and the families with Caucasian ethnicity ($n = 133$, HLOD = 1.22, $p = 0.018$).

Family-based and population-based association tests

The family-based association study in 159 HPC families did not provide evidence for overtransmission of either the A1 or A2 allele to offspring with prostate cancer and neither did the alleles at 2 microsatellite markers D10S192 and D10S222. The χ^2 were 1.36 (df = 1, $p = 0.24$) for CYP17, 5.13 (df = 9, $p = 0.82$) for D10S192 and 4.52 (df = 6, $p = 0.61$) for D10S222. To decrease the impact of different racial groups in the sample as a possible confounder, the family-based association tests were reperformed in the 133 Caucasian families. No statistically significant overtransmission of any allele in CYP17 or the 2 microsatellite markers was found.

The CYP17 A1/A2 polymorphism was also genotyped in the 249 sporadic cases and in the 211 unaffected controls. The SNP was in HWE in each subset. The allele and genotype frequencies were compared among HPC probands, sporadic cases and unaffected controls. To decrease the confounding factor of racial differences, the comparison was limited to Caucasians only. The allele frequency of the A2 allele was 43% in 133 HPC probands, 39% in 225 sporadic cases and 36% in 182 unaffected controls. The differences in allele frequencies were not statistically significant ($p = 0.23$ between HPC probands and controls, $p = 0.55$ between sporadic cases and controls and $p = 0.34$ between all cases and controls). No statistically significant differences were observed in genotype frequencies between HPC probands, sporadic prostate cancer patients, and unaffected control subjects, although a trend toward more A2 heterozygotes and homozygotes in prostate cancer patients, especially in the HPC probands, was observed (Table II). For example, using A1 homozygotes as a reference group, the point estimate of relative risk for HPC in individuals who are homozygous for A2 was 1.62 (95% confidence interval [CI] = 0.83–3.18), for sporadic prostate cancer was 1.04 (95% CI = 0.57–1.91) and for all prostate cancer was 1.25 (95% CI = 0.72–2.15).

Since the evidence for linkage at the CYP17 gene region in our study was mainly from the families with older mean age of diagnosis and the associations between the CYP17 A1/A2 polymorphism and prostate cancer risk were observed in age-stratified subgroups in other studies, the possibility that the influence of the CYP17 gene is age-dependent was explored next. The study subjects were stratified into 2 age groups (≤60 and >60 years), and the genotypes were then compared between HPC probands, sporadic cases and controls within the same age groups. The differences in allele and genotype frequencies were not statistically significant between HPC probands, sporadic prostate cancer patients and unaffected control subjects in either group (Table II).

The relationship of the A1/A2 polymorphism with Gleason scores or pathologic stages in sporadic prostate cancer cases was examined. No statistically significant difference in the genotypic frequencies of the SNP was found between the groups with low (≤6) and high (≥7) Gleason scores or between the group with

TABLE I – MULTIPOINT LINKAGE RESULTS IN 159 HPC FAMILIES

	No. of pedigrees	HLOD			Allele sharing LOD		
		D10S192	CYP17	D10S222	D10S192	CYP17	D10S222
All families	159	0.62	1.00	1.32	0.33	0.35	0.46
Age at diagnosis							
<65	79	0.17	0.22	0.28	0.14	0.09	0.10
≥65	80	0.53	0.93	1.21	0.18	0.29	0.43
Number of affected members							
3	29	0.00	0.00	0.00	0.00	0.00	0.00
4	40	0.26	0.25	0.35	0.09	0.08	0.13
≥5	90	0.43	0.88	1.13	0.59	0.78	0.91
Ethnicity							
Caucasian	133	0.37	0.87	1.22	0.33	0.48	0.65
African-American	14	0.24	0.19	0.19	0.01	0.00	0.00
Other	12	0.11	0.00	0.00	0.00	0.00	0.00

TABLE II – CYP17 A1/A2 GENOTYPE FREQUENCIES FOR HPC PROBANDS, SPORADIC CASES AND UNAFFECTED CONTROL SUBJECTS

CYP17	Frequency (%)			OR[1,2] (95% CI)	OR[3] (95% CI)	OR[4] (95% CI)
	HPC proband	Sporadic	Control			
All patients	(n = 133)	(n = 225)	(n = 182)			
A1/A1	0.33	0.39	0.42	1	1	1
A1/A2	0.49	0.47	0.43	1.46 (0.88–2.43)	1.18 (0.77–1.80)	1.27 (0.86–1.88)
A2/A2	0.18	0.14	0.14	1.62 (0.83–3.18)	1.04 (0.57–1.91)	1.25 (0.72–2.15)
Any A2	0.67	0.61	0.57	1.49 (0.93–2.40)	1.14 (0.77–1.70)	1.27 (0.88–1.82)
Patients >60 years	(n = 85)	(n = 96)	(n = 71)			
A1/A1	0.32	0.43	0.45	1	1	1
A1/A2	0.5	0.47	0.34	2.05 (0.99–4.20)	1.61 (0.80–3.24)	1.76 (0.94–3.27)
A2/A2	0.18	0.1	0.21	1.19 (0.49–2.86)	0.56 (0.22–1.44)	0.78 (0.36–1.69)
Any A2	0.68	0.57	0.55	1.74 (0.90–3.34)	1.21 (0.64–2.30)	1.39 (0.79–2.43)
Patients ≤60 years	(n = 48)	(n = 129)	(n = 111)			
A1/A1	0.33	0.36	0.4	1	1	1
A1/A2	0.48	0.47	0.5	1.20 (0.56–2.56)	1.05 (0.61–1.82)	1.09 (0.65–1.81)
A2/A2	0.19	0.16	0.1	2.29 (0.80–6.45)	1.91 (0.82–4.45)	2.02 (0.91–4.48)
Any A2	0.67	0.64	0.57	1.41 (0.69–2.90)	1.19 (0.70–2.00)	1.23 (0.76–2.01)

[1]All ORs were age adjusted.–[2]HPC probands *vs.* controls.–[3]Sporadic cases *vs.* controls.–[4]All cases *vs.* controls.

TABLE III – CYP17 A1/A2 GENOTYPE FREQUENCIES IN SPORADIC CASES (CAUCASIANS ONLY)

CYP17	Gleason score, n (%)		Pathologic stage, n (%)	
	≤6	≥7	0	≥1
A1/A1	29 (34.12)	59 (42.14)	29 (41.43)	59 (38.06)
A1/A2	42 (49.41)	64 (45.71)	30 (42.86)	76 (49.03)
A2/A2	14 (16.47)	17 (12.14)	11 (15.71)	20 (12.90)

disease confined to the prostate compared to the group with non-localized disease (Table III).

DISCUSSION

Although there have been several reports on the association between the 5′ promoter polymorphism of CYP17 gene and risk for prostate cancer using population-based association study design, the results are inconclusive.[30,31,33,34] Considering the importance of the CYP17 gene and androgen pathway in the etiology, prevention and treatment of prostate cancer, additional studies to evaluate this polymorphism and its association with prostate cancer are clearly warranted. Furthermore, 2 other important questions have not been addressed previously: *(i)* What is the possible association with hereditary prostate cancer; *(ii)* what evidence exists for the CYP17 gene as a prostate cancer susceptibility gene using the genetic linkage approach. Our study was designed to address these important areas. We found suggestive evidence for prostate cancer linkage to the region of 10q24.3, which contains the CYP17 gene in the total 159 HPC families. However, we did not observe a statistically increased risk to sporadic prostate cancer or to hereditary prostate cancer in subjects with the A2 variant of the 5′ promoter polymorphism in the CYP17 gene.

A genetic linkage study is one important study design to evaluate a candidate gene in complex diseases such as prostate cancer. A significant feature of linkage analysis is that it is insensitive to allelic heterogeneity. If a mutation has a large effect (*i.e.*, high penetrance) and there are multiple such mutations within a gene, a linkage study is likely to detect such a gene while family-based or population-based association approaches are likely to fail. Information regarding specific sequence variants within a gene is not necessary for a linkage study, but this is essential for association studies.

The HLOD of 1.30 (p = 0.014) from our study did not reach the criteria for suggestive evidence for linkage from a genome-wide screen.[47] However, the linkage evidence warranted further evaluation for the following reasons: *(i)* Our linkage study was clearly not a genome-wide screen and we had a very specific hypothesis, *i.e.*, whether there is linkage at the CYP17 region. Thus, the genome-wide criteria are too stringent in our case. *(ii)* We did not use multiple genetic models. Instead, we used only 1 model that was defined previously.[8] *(iii)* The empirical p-value of the observed HLOD of 1.3 was 0.008 and is similar to the χ^2-based p-value that we reported. The empirical p-value was obtained by simulating and analyzing 10,000 replicates generated assuming no linkage between a prostate cancer susceptibility gene and an 8-allele marker using the exactly same pedigree structure and the availability of genotypes in the 159 HPC families.

It is interesting that the linkage evidence for 10q24.3 region was stronger in the parametric analyses than in the nonparametric analyses. Several factors might contribute to this observation. First, parametric analyses are likely to perform better than nonparametric analysis when there are substantial numbers of phenocopies, which is likely in prostate cancer. Phenocopies are incorporated into the genetic model in the parametric analyses, whereas all affecteds are considered as disease gene carriers in the nonparametric analysis. Second, parametric analyses generally have more power than nonparametric analyses when the specified genetic models are close to the true model.[48,49] The autosomal dominant model assumed in the analysis is consistent with the results of 4 segregation studies.[2,4,6,7] If the underlying model was recessive, we would expect to observe stronger linkage results from the nonparametric analysis because it is based on an allele sharing method.

Two other genome-wide linkage studies also showed moderate evidence for linkage to chromosomal 10q.[12,50] However, the linkage signals in both studies were at 10q25–qter, which is approximately 30 cM telomeric to the CYP17 locus where we observed evidence of linkage. Interestingly, loss of heterozygosity (LOH) in the region of 10q23–25, which includes the region with evidence of linkage in our study, was found to be a frequent event in prostate carcinoma (approximately 50% of tumors studied), as well as in other tumors including glioblastoma multiforme, endometrial carcinoma, breast carcinoma and melanoma.[51–57] A tumor-suppressor gene(s) located in this chromosomal region, including PTEN, was proposed to be involved in the development of tumors.[58–60] It is also possible that the linkage signal we observed for the CYP17 locus is due to the nearby tumor-suppressor gene. Further study in and near this chromosomal region is necessary to clarify the possibility.

A family-based association study has its advantages in detection of a mutation with a large effect on disease and in the minimization of false-positive findings due to racial admixture in study samples.[61] However, the power to detect a mutation is severely impaired when multiple mutations (*i.e.*, allelic heterogeneity) exist[62] or when a mutation has a small effect.[61] There are at least 3

TABLE IV – REPORTED ASSOCIATION STUDIES BETWEEN CYP17 GENOTYPES AND RISK FOR PROSTATE CANCER

Study population	No. of subjects	OR (95% CI)	Age-stratified OR (95% CI)	References
Caucasians	96 cases 159 controls	For A1/A2 + A2/A2 1.7 (1.0–3.1)	≤64 years 2.3 (1.0–4.8)	Lunn *et al.*[30]
Australians	63 cases 126 controls	For A2/A2 2.8 (1.02–7.76)	>66 years 8.93 (1.78–49.19)	Gsur *et al.*[31]
Caucasians	178 cases 160 controls	For A1/A1 1.61 (1.02–2.53)	NA	Wadelius *et al.*[33]
Japanese	252 cases 131 controls	For A1/A1 2.57 (1.39–4.78)	≥73 years 2.95 (1.32–6.63)	Habuchi *et al.*[34]
Japanese	105 cases 210 controls	For A2/A2 2.39 (1.04–5.46)	<72 years 4.09 (1.05–15.9)	Yamada *et al.*[32]
Caucasians	225 cases 133 HPC probands 182 controls	For A1/A2 + A2/A2 1.27 (0.88–1.82)		Present study

explanations for our negative findings using the family-based association test: *(i)* the polymorphism does not increase the risk for prostate cancer; *(ii)* the sequence variant increases the risk but there are other mutations (*i.e.*, allelic heterogeneity) and our study does not have adequate power to detect such a risk; *(iii)* the sequence variant only increases the risk slightly (*i.e.*, low penetrance) and thus does not completely segregate with prostate cancer in these families.

A population-based association study is an alternative study design to detect sequence variants with low penetrance. However, our study failed to observe a statistically significant increased frequency of A2 carriers in either HPC or sporadic prostate cancer patients. Our study also failed to observe statistically significant increased frequency of A2 carriers in early age of onset prostate cancer cases. Caution should be exercised to interpret these results for 2 reasons. First, our study has reasonable but limited power to detect a mutation with a small effect. Using a point estimate of relative risk of 1.7 for sporadic prostate cancer,[30] a carrier rate of 57% in control subjects and a significance level of 0.05 (2-tailed test), our sample has 72% or 79% power to detect such a mutation, respectively, in 225 sporadic prostate cancer patients and in all 355 prostate cancer patients. Second, there is potential misclassification in our control subjects. Although these subjects had normal digital rectal examination (DRE) and normal PSA level (*i.e.*, <4 ng/ml), some of them are young enough that they could be disease gene carriers and develop prostate cancer later. The ORs for the polymorphism in our study were adjusted for age. This may alleviate the problem but cannot totally remove the confounder.

The lack of association between the CYP17 A1/A2 polymorphism and the risk for prostate cancer in our study was in accordance with the *in vitro* electromobility shift assay conducted by Nedelcheva Kristensen *et al.*[63] They observed no binding of human Sp-1 recombinant protein to either A1 or A2 allele and concluded that the T to C transition polymorphism does not create an Sp-1 binding site and has no effect on the expression of the CYP17 gene. A study by Allen *et al.*[64] also showed no association between the CYP17 A1/A2 polymorphism and serum testosterone levels. These findings support the possibility that the CYP17 A1/A2 polymorphism may have no effect on the expression of the CYP17 gene and hence no effect on androgen levels or risk of prostate cancer.

By reviewing the published studies evaluating the 5′ promoter polymorphism of the CYP17 gene and prostate cancer, several common features can be found (Table IV). First, most of the studies were based on relatively small numbers of subjects. Second, the point estimates for the relative risks were small and 95% CI were large, ranging from 1.61 (95% CI = 1.02–2.53) in the study by Wadelius *et al.*[33] to 2.8 (95% CI = 1.78–49.19) in the study by Gsur *et al.*[31] Third, different alleles increased risk for prostate cancer in different studies. While the A2 allele increased risk in the studies carried out in the United States,[30] Austria[31] and Japan,[32] the A1 allele increased prostate cancer risk in the studies carried out in Sweden[33] and Japan.[34] Although different genetic backgrounds and also environmental factors may contribute to this difference, the results also suggest that the 5′ promoter polymorphism may not be causal, but might instead be in linkage disequilibrium with a disease causal mutation within the CYP17 gene. There have been 12 more SNPs identified within the CYP17 gene (http://www.ncbi.nlm.nih.gov/SNP/snp_ref.cgi?locusId=1586). The association between these SNPs and the risk for prostate cancer needs to be evaluated.

In conclusion, future studies with large numbers of prostate cancer patients (hereditary and sporadic) and well characterized unaffected control subjects are needed to clarify whether this polymorphism increases the risk for prostate cancer. Evaluating other polymorphisms in the exons and introns within the gene is crucial. Furthermore, considering the multiple enzymatic steps involved in androgen biosynthesis and metabolism, a pathway-wide study investigating multiple SNPs in multiple genes simultaneously may allow for a more precise estimation of inherited risk for prostate cancer.

ACKNOWLEDGEMENTS

The authors thank all of the study subjects who participated in our study. This work was partially supported by PHS SPORE and 2 grants from the Department of Defense to W.B.I and J.X. We thank Dr. T. Howard for his advice and excellent technical support. We thank Mr. J. Clark, Ms. B. Fitzerrald, Ms. J. Haynes and Ms. M. Jones for their excellent technical support.

REFERENCES

1. Steinberg GD, Carter BS, Beaty TH, et al. Family history and the risk of prostate cancer. Prostate 1990;17:337–47.
2. Carter BS, Beaty TH, Steinberg GD. Mendelian inheritance of familial prostate cancer. Proc Natl Acad Sci USA 1992;89:3367–71.
3. Gronberg H, Damber L, Damber JE. Studies of genetic factors in prostate cancer in a twin population. J Urol 1994:152(5 Pt 1):1484–7.
4. Gronberg H, Damber L, Damber JE, et al. Segregation analysis of prostate cancer in Sweden: support for dominant inheritance. Am J Epidemiol 1997;146:552–7.
5. Page WF, Braun MM, Partin AW, et al. Hereditary and prostate cancer: a study of World War II veteran twins. Prostate 1997;33:240–5.
6. Schaid DJ, McDonnell SK, Blute ML, et al. Evidence for autosomal dominant inheritance of prostate cancer. Am J Hum Genet 1998;62: 1425–38.
7. Cui J, Staples MP, Hopper JL, et al. Segregation analyses of 1,476 population-based Australian families affected by prostate cancer. Am J Hum Genet 2001;68:1207–18.
8. Smith JR, Freije D, Carpten JD, et al. Major susceptibility locus for prostate cancer on chromosome 1 suggested by a genome-wide search. Science 1996;274:1371–4.
9. Xu J, ICPCG. Combined analysis of hereditary prostate cancer linkage to 1q24–25: results from 772 hereditary prostate cancer families

from the International Consortium for Prostate Cancer Genetics. Am J Hum Genet 2000;66:945–57.
10. Berthon P, Valeri A, Cohen-Akenine A, et al. Predisposing gene for early-onset prostate cancer, localized on chromosome 1q42.2–43. Am J Hum Genet 1998;62:1416–24.
11. Xu J, Meyers DA, Freije D, et al. Evidence for a prostate cancer susceptibility locus on the X chromosome. Nat Genet 1998;20:175–9.
12. Gibbs M, Stanford JL, Jarvik GP, et al. A genomic scan of families with prostate cancer identifies multiple regions of interest. Am J Hum Genet 2000;67:100–9.
13. Berry R, Schaid DJ, Smith JR, et al. Linkage analyses at the chromosome 1 loci 1q24–25 (HPC1), 1q42.2–43 (PCAP), and 1p36 (CAPB) in families with hereditary. Am J Hum Genet 2000;66:539–46.
14. Ross RK, Pike MC, Coetzee GA, et al. Androgen metabolism and prostate cancer: establishing a model of genetic susceptibility. Cancer Res 1998;58:4497–504.
15. Ruijter E, Kaa CVD, Miller G, et al. Molecular genetics and epidemiology of prostate carcinoma. Endocrine Rev 1999;20:22–45.
16. Henderson BE, Feigelson HS. Hormonal carcinogenesis. Carcinogenesis 2000;21(3):427–33.
17. Gann PH, Hennekens CH, Ma J, et al. Prospective study of sex hormone levels and risk of prostate cancer. J Natl Cancer Inst 1996; 88:1118–26.
18. Ross RK, Schottenfeld D. Cancer epidemiology and prevention. New York: Oxford University Press, 1996. 1180–206.
19. Picado-Leonard J, Miller WL. Cloning and sequence of the human gene for P450c17 (steroid 17 α-hydroxylase/17,20 lyase): similarity with the gene for P450c21. DNA 1987;6:439–48.
20. Carey AH, Waterworth D, Patel K, et al. Polycystic ovaries and premature male pattern baldness are associated with one allele of the steroid metabolism gene CYP17. Hum Mol Genet 1994;3:1873–6.
21. Feigelson HS, Shames LS, Pike MC, et al. Cytochrome P450c17α gene (CYP17) polymorphism is associated with serum estrogen and progesterone concentrations. Cancer Res 1998;58:585–7.
22. Haiman CA, Hankinson SE, Spiegelman D, et al. The relationship between the polymorphism in CYP17 gene with plasm hormone levels and breast cancer. Cancer Res 1999;59:1015–20.
23. Feigelson HS, Coetzee GA, Kolonel LN, et al. A polymorphism in the CYP17 gene increase the risk fo breast cancer. Cancer Res.1997;57: 1063–5.
24. Bergman-Jungestrom M, Gentile M, Lundin AC, et al. Association between CYP17 gene polymorphism and risk of breast cancer in young women. Int J Cancer 1999;84:350–3.
25. Young IE, Kurian KM, Annink C, et al. A polymorphism in the CYP17 gene is associated with male breast cancer. Br J Cancer 1999;81:141–3.
26. Dunning AM, Healey CS, Pharoah PD, et al. No association between a polymorphism in the steroid metabolism gene CYP17 and the risk of breast cancer. Br J Cancer 1998;77:2045–7.
27. Helzlsouer KJ, Huang HY, Strickland PT, et al. Association between CYP17 polymorphisms and the development of breast cancer. Cancer Epidemiol Biomark Prev 1998;7:945–9.
28. Techatraisak K, Conway GS, Rumsby G. Frequencies of the polymorphism in the regulatory region of the 17 α-hydroxylase-17,20-lyase (CYP17) gene in hyperandrogenic states. Clin Endocrinol 1997; 46:131–4.
29. Weston A, Pan CF, Bleiweiss IJ, et al. CYP17 genotype and breast cancer risk. Cancer Epidemiol Biomark Prev 1998;7:941–4.
30. Lunn RM, Bell DA, Mohler JL, et al. Prostate cancer risk and polymorphism in 17 hydroxylase (CYP17) and steroid reductase (SRD5A2). Carcinogenesis 1999;20:1727–31.
31. Gsur A, Bernhofer G, Hinteregger S, et al. A polymorphism in the CYP17 gene is associated with prostate cancer risk. Int J Cancer 2000;87:434–7.
32. Yamada Y, Watanabe M, Murata M, et al. Impact of genetic polymorphisms of 17-hydroxylase cytochrome P-450 (CYP17) and steroid 5alpha-reductase type II (SRD5A2) genes on prostate-cancer risk among the Japanese population. Int J Cancer 2001;92:683–6.
33. Wadelius M, Anderson AO, Johanson JE, et al. Prostate cancer associated with CYP17 genotype. Pharmacogenetics 1999;9:635–9.
34. Habuchi T, Liqing Z, Suzuki T, et al. Increased risk of prostate cancer and benign prostate hyperplasia associated with a CYP17 gene polymorphism with a gene dosage effect. Cancer Res 2000;60:5710–3.
35. Xu J, Zheng SL, Carpten JD, et al. Evaluation of linkage and association of HPC2/ELAC2 in patients with familial or sporadic prostate cancer. Am J Hum Genet 2001;68:901–11.
36. Broman KW, Murray JC, Sheffield VC, et al. Comprehensive human genetic maps: individual and sex-specific variation in recombination. Am J Hum Genet 1998;63:861–9.
37. Collins A, Teague J, Keats BJ, et al. Linkage map integration. Genomics 1996;36:157–62.
38. Lathrop GM, Lalouel JM, Julier C, et al. Strategies for multilocus linkage analysis in humans. Proc Natl Acad Sci USA 1984;81:3443–6.
39. Cottingham RW Jr., Idury RM, Schaffer AA. Faster sequential genetic linkage computations. Am J Hum Genet 1993;53:252–63.
40. Weir BS. Genetic data analysis II: methods for discrete population genetic data. Sunderland, MA: Sinauer Association, Inc. Publishers, 1996.
41. Kruglyak L, Daly MJ, Reeve-Daly MP, et al. Parametric and nonparametric linkage analysis: a unified multipoint approach. Am J Hum Genet 1996;58:1347–63.
42. Kong A, Cox NJ. Allele-sharing models: LOD scores and accurate linkage tests. Am J Hum Genet 1997;61:1179–88.
43. Ott J. Analysis of human genetic linkage. 3rd ed. Baltimore: Johns Hopkins Press, 1998.
44. Whittemore A, Halpern J. A class of tests for linkage using affected pedigree members. Biometrics 1994;50:118–27.
45. Faraway JJ. Distribution of the admixture test for the detection of linkage under heterogeneity. Genet Epidemiol 1993;10:75–83.
46. Laird NM, Horvath S, Xu X. Implementing a unified approach to family based test of association. Genet Epidemiol 2000;19 (Suppl 1):S36–S42.
47. Lander ES, Kruglyak L. Genetic dissection of complex traits: guideline for interpreting and reporting linkage results. Nat Genet 1995;11: 241–7.
48. Greenberg DA, Abreu P, Hodge SE. The power to detect linkage in complex disease by means of simple LOD-score analyses. Am J Hum Genet 1998;63:870–9.
49. Durner M, Vieland VJ, Greenberg DA. Further evidence for the increased power of LOD scores compared with nonparametric methods. Am J Hum Genet 1999;64:281–9.
50. Suarez BK, Lin J, Burmester JK, et al. A genome screen of multiplex sibships with prostate cancer. Am J Hum Genet 2000;66:933–44.
51. Isaacs WB, Carter BS. Genetic changes associated with prostate cancer in humans. Cancer Surv. 1991;11:15–24.
52. Gray IC, Phillips SM, Lee SJ, et al. Loss of the chromosomal region 10q23–25 in prostate cancer. Cancer Res 1995;55:4800–3.
53. Ittmann M. Allelic loss on chromosome 10 in prostate adenocarcinoma. Cancer Res 1996;56:2143–7.
54. Komiya A, Suzuki H, Ueda T, et al. Allelic losses at loci on chromosome 10 are associated with metastasis and progression of human prostate cancer. Genes Chromosomes Cancer 1996;17:245–53.
55. Trybus TM, Burgess AC, Wojno KJ, et al. Distinct areas of allelic loss on chromosomal regions 10p and 10q in human prostate cancer. Cancer Res 1996;56:2263–7.
56. Bose S, Wang SI, Terry MB, et al. Allelic loss of chromosome 10q23 is associated with tumor progression in breast carcinomas. Oncogene 1998;17:123–7.
57. Rasheed BK, Wiltshire RN, Bigner SH, et al. Molecular pathogenesis of malignant gliomas. Curr Opin Oncol 1999;11:162–7.
58. Li DM, Sun H. TEP1, encoded by a candidate tumor suppressor locus, is a novel protein tyrosine phosphatase regulated by transforming growth factor beta. Cancer Res 1997;57:2124–9.
59. Li J, Yen C, Liaw D, et al. PTEN, a putative protein tyrosine phosphatase gene mutated in human brain, breast, and prostate cancer. Science 1997;275:1943–7.
60. Steck PA, Pershouse MA, Jasser SA, et al. Identification of a candidate tumour suppressor gene, MMAC1, at chromosome 10q23.3 that is mutated in multiple advanced cancers. Nat Genet 1997;15:356–62.
61. Risch N, Merikangas K. The future of genetic studies of complex human diseases. Science 1996;273:1516–7.
62. Slager SL, Huang J, Vieland VJ. Effect of allelic heterogeneity on the power of the transmission disequilibrium test. Genet Epidemiol 2000; 18:143–56.
63. Nedelcheva Kristensen V, Haraldsen EK, Anderson KB, et al. CYP17 and breast cancer risk: the polymorphism in the 5′ flanking area of the gene does not influence binding to Sp-1. Cancer Res 1999;59:2825–8.
64. Allen NE, Forrest MS, Key TJ. The association between polymorphisms in the CYP17 and 5alpha-reductase (SRD5A2) genes and serum androgen concentrations in men. Cancer Epidemiol Biomark Prev 2001;10:185–9.

[CANCER RESEARCH 62, 1784–1789, March 15, 2002]

Joint Effect of *HSD3B1* and *HSD3B2* Genes Is Associated with Hereditary and Sporadic Prostate Cancer Susceptibility[1]

Bao-li Chang, Siqun L. Zheng, Gregory A. Hawkins, Sarah D. Isaacs, Kathy E. Wiley, Aubrey Turner, John D. Carpten, Eugene R. Bleecker, Patrick C. Walsh, Jeffrey M. Trent, Deborah A. Meyers, William B. Isaacs, and Jianfeng Xu[2]

Center for Human Genomics, Wake Forest University School of Medicine, Winston-Salem, North Carolina 27157 [B-l. C., S. L. Z., G. A. H., A. T., E. R. B., D. A. M., J. X.]; University of Maryland School of Medicine, Baltimore, Maryland 21201 [B-l. C.]; Department of Urology, Johns Hopkins Medical Institutions, Baltimore, Maryland 21287 [S. D. I., K. E. W., P. C. W., W. B. I.]; National Human Genome Research Institute, NIH, Bethesda, Maryland 20892 [J. D. C., J. M. T.]

ABSTRACT

3β-hydroxysteroid dehydrogenases (HSD3Bs), encoded by the *HSD3B* gene family at 1p13, have long been hypothesized to have a major role in prostate cancer susceptibility. The recent reports of a prostate cancer linkage at *1p13* provided additional evidence that *HSD3B* genes may be prostate cancer susceptibility genes. To evaluate the possible role of *HSD3B* genes in prostate cancer, we screened a panel of DNA samples collected from 96 men with or without prostate cancer for sequence variants in the putative promoter region, exons, exon-intron junctions, and 3′-untranslated region of *HSD3B1* and *HSD3B2* genes by direct sequencing. Eleven single nucleotide polymorphisms (SNPs) were identified, four of which, including a missense change (*B1-N367T*), were informative. These four SNPs were further genotyped in a total of 159 hereditary prostate cancer probands, 245 sporadic prostate cancer cases, and 222 unaffected controls. Although a weak association between prostate cancer risk and a missense SNP (*B1-N367T*) was found, stronger evidence for association was found when the joint effect of the two genes was considered. Men with the variant genotypes at either *B1-N367T* or *B2-c7519g* had a significantly higher risk to develop prostate cancer, especially the hereditary type of prostate cancer. Most importantly, the subset of hereditary prostate cancer probands, whose families provided evidence for linkage at 1p13, predominantly contributed to the observed association. Additional studies are warranted to confirm these findings.

INTRODUCTION

Prostate cancer (MIM 176807) is the most frequently diagnosed cancer and the second leading cause of cancer mortality among men in many industrialized countries. Evidence for genetic susceptibility to prostate cancer is well documented from epidemiological studies (1), twin studies (2–4), and segregation analyses (5–8). Chromosomal regions that are likely to contain prostate cancer susceptibility genes have been identified including *HPC1*[3] at 1q24–25 (9), *PCAP* at 1q42–43 (10), *HPCX* at Xq27–28 (11), *CAPB* at 1p36 (12), *HPC20* at 20q13 (13), *HPC2* at 17p11 (14, 15), and 8p22–23 (16).

Androgens have been hypothesized to be involved in prostate carcinogenesis because of their essential role in prostate development, growth, and maintenance. The enzyme HSD3B is a critical component of the androgen metabolism pathway because it catalyzes androstenedione production in steroidogenic tissues and converts the active dihydrotestosterone into inactive metabolites in steroid target tissues. The *HSD3B* gene family has two genes and five pseudogenes, all of which map to chromosome 1p13 (17–19). The *HSD3B1* gene encodes the type I enzyme, which is exclusively expressed in the placenta and peripheral tissues, such as prostate, breast, and skin. The *HSD3B2* gene encodes the type II enzyme, which is predominantly expressed in classical steroidogenic tissues, namely the adrenals, testis, and ovary (18, 20–23). A number of mutations in *HSD3B2* has been found to cause congenital adrenal hyperplasia, a rare Mendelian disease, manifested by salt-wasting and incomplete masculinization in males (24).

Recent linkage findings at 1p13 significantly increase the likelihood that *HSD3B* genes play an important role in prostate cancer susceptibility. In a chromosome-wide linkage study to evaluate different prostate cancer susceptibility loci on chromosome 1 in 159 HPC families, our group reported evidence for linkage in a broad region from 1p13 to 1q32 (25). The LOD score assuming heterogeneity was 1.31 ($P = 0.01$), and the allele-sharing LOD score was 1.34 ($P = 0.01$) at *HSD3B2*. The evidence for linkage was stronger in families with five or more affected men (allele-sharing LOD = 2.22, $P = 0.001$) and in families with mean age of onset > 65 years (allele-sharing LOD = 1.45, $P = 0.01$). In another genome-wide scan for prostate cancer susceptibility loci, Goddard *et al.* (26) reported a LOD score of 3.25 ($P = 0.0001$) at 1p13, near markers *D1S534* and *D1S1653*, when the Gleason score was included as a covariate.

There are only a few studies on the sequence variants of *HSD3B2* in prostate cancer. A complex $(TG)_n$ $(TA)_n$ $(CA)_n$ repeat has been described and studied in intron 3 of *HSD3B2* (27, 28). However, there is no published study that evaluates the association between this repeat and other sequence variants in *HSD3B1* and prostate cancer risk. Considering the biological importance of the *HSD3B* genes and the evidence that these genes are located in a chromosomal region that is likely to contain prostate cancer susceptibility genes, a systematic study and evaluation of these genes in relationship to prostate cancer appears warranted.

We have two major goals in this study. The first one is to identify sequence variants in the *HSD3B1* and *HSD3B2* genes by directly sequencing the PCR products from the 500-bp promoter region, all exons, exon-intron junctions, and 3′-UTR of both genes in 96 subjects. The second goal is to test for association between prostate cancer and *HSD3B* genes by comparing the distributions of the four frequent SNPs in 159 HPC probands, 245 sporadic prostate cancer cases, and 222 unaffected controls.

MATERIALS AND METHODS

Subjects. A detailed description of the study sample was presented elsewhere (25). HPC probands ($n = 159$) were ascertained at the Brady Urology Institute at Johns Hopkins Hospital (Baltimore, MD) through referrals, review of medical records for patients seen at Johns Hopkins Hospital for treatment of prostate cancer, and respondents to various lay publications describing our studies. Each proband had at least two first-degree relatives affected with prostate cancer. The diagnosis of prostate cancer was verified by medical records. The mean age at prostate cancer diagnosis for these probands was 61 years; 133 (84%) were Caucasian, and 14 (8.8%) were African-American.

Received 8/31/01; accepted 1/9/02.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[1] Supported by PHS SPORE CA58236 and two grants from the Department of Defense (to W. B. I. and J. X.).

[2] To whom requests for reprints should be addressed, at Center for Human Genomics, Wake Forest University School of Medicine, Medical Center Boulevard, Winston-Salem, NC 27157. Phone: (336) 716-5700; Fax: (336) 716-7575; E-mail: jxu@wfubmc.edu.

[3] The abbreviations used are: HPC, hereditary prostate cancer; HSD3B, 3β-hydroxysteroid dehydrogenase; SNP, single nucleotide polymorphism; LOD, log of odd; UTR, untranslated region; PSA, prostate specific antigen; HWE, Hardy-Weinberg equilibrium; LD, linkage disequilibrium; RR, relative risk; CI, confidence interval; PKC, protein kinase C; IL, interleukin.

All 245 unrelated prostate cancer cases were recruited from patients who underwent treatment for prostate cancer at the Johns Hopkins Hospital and did not have first-degree relatives affected with prostate cancer. For each subject, the diagnosis of prostate cancer was confirmed by pathology reports. Preoperative PSA levels, Gleason score, and pathological stages were available for 202, 240, and 241 cases, respectively. Mean age at diagnosis for these cases was 58.7 years. Over 93% of the cases were Caucasian, and 3.2% were African-American.

Nonprostate cancer controls (222) were selected from men participating in screening programs for prostate cancer. By applying the exclusion criteria of abnormal digital rectal examination and abnormal PSA level (*i.e.*, ≥4 ng/ml), 211 were eligible for the study. The mean age at examination was 58 years. Over 86% of the eligible controls were Caucasian, and 7.1% were African-American. On the basis of interview of the subjects, 5.6% of the eligible controls had brothers or their father affected with prostate cancer.

The Institutional Review Board of Johns Hopkins University approved the protocols for subject recruitment. After each participant was guided through an informed consent process, they completed a signed consent form as a record of this process.

Sequencing Methods and SNP Genotyping. The *HSD3B1* and *HSD3B2* genes are structurally very similar, with 85% homology (17, 20, 29, 30). Both genes span 7.8 kb and contain 4 exons. To identify SNPs in *HSD3B1* and *HSD3B2*, we directly sequenced the PCR products of the putative promoter region, all exons, exon-intron junctions, and the 3′-UTR of both genes in 96 subjects. These subjects include 72 Caucasians and 24 African-Americans, with equal numbers of HPC cases, sporadic cases, and unaffected controls in each racial group. Table 1 lists the primers used to amplify the PCR products, the sizes of amplified PCR fragments, and the annealing temperatures for each pair of primers. All PCR reactions were performed in a 30-μl volume consisting of 30 ng of genomic DNA, 0.2 μM each primer, 0.2 mM each deoxynucleotide triphosphate, 1.5 mM $MgCl_2$, 20 mM Tris-HCl, 50 mM KCl, and 0.5 units of Taq polymerase (Life Technologies, Inc.). PCR cycling conditions were as follows: a 94°C hotstart for 4 min, followed by 33 cycles of 94°C for 30 s, specified annealing temperature for 30 s, and 72°C for 30 s, with a final extension of 72°C for 6 min. All PCR products were purified using the QuickStep PCR purification Kit (Edge BioSystems, Gaithersburg, MD) to remove deoxynucleotide triphosphates and excess primers. All sequencing reactions were performed using dye-terminator chemistry (BigDye; ABI, Foster City, CA) and then precipitated using 63 +/− 5% ethanol. Samples were loaded onto an ABI 3700 DNA Analyzer after adding 10 μl of formamide. SNPs were identified using Sequencher software version 4.0.5 (Gene Codes Corp.). For the four frequent SNPs, additional genotyping of 159 HPC probands, 245 sporadic prostate cancer cases, and 222 unaffected controls was performed using the same sequencing method.

Statistical Methods. HWE tests for all SNPs, and LD tests for all pairs of SNPs, were performed using the Genetic Data Analysis (GDA) computer program (31). The HWE tests were based on exact tests, where a large number of the possible arrays were generated by permuting the alleles among genotypes, and the proportion of these permuted genotypic arrays with a smaller conditional probability than the original data was calculated. The LD tests were based on an exact test assuming multinomial probability of the multilocus genotype, conditional on the single-locus genotype (32). A Monte Carlo simulation was used to assess the significance by permuting the single-locus genotypes among individuals in the sample to simulate the null distribution. The empirical *P*s of both HWE and LD tests were based on 10,000 replicate samples.

Association tests between the SNPs and prostate cancer were performed by comparing allele and genotype frequencies between cases and controls for each SNP. Allele frequencies were estimated by a direct count. The hypotheses of differences in allele frequencies between cases and controls were tested using standard contingency χ^2 tests, and *P*s were determined via χ^2 approximation (33). Differences in genotype frequencies (variant alleles were assumed to be dominant or recessive) between cases and controls were tested using unconditional logistic regression and were adjusted for potential confounders, such as age.

RESULTS

SNP Identification. A total of five SNPs in *HSD3B1* and six SNPs in *HSD3B2* were identified in the screening panel of 96 subjects. The frequency of the SNPs by race and prostate cancer status are presented in Table 2. There were four SNPs in the coding region of *HSD3B1* (exon 4) and two of which are nonsynonymous changes. SNP *B1-F286L* causes an amino acid change from phenylalanine to leucine, and *B1-N367T* results in an amino acid change from asparagine to threonine. The possible effects of these two SNPs on the functional enzymatic activities of *HSD3B1* protein remain to be determined. Although no SNPs were identified in the coding region of *HSD3B2*, two common SNPs (*B2-c7474t* and *B2-c7519g*) were found in the 3′-UTR region.

Association between Prostate Cancer Susceptibility and the SNPs. The four frequent SNPs (two each in *HSD3B1* and *HSD3B2*) observed in the panel of 96 subjects were further studied in a larger study population of 159 HPC probands, 245 sporadic prostate cancer cases, and 222 unaffected controls. To decrease the potential impact of population stratification, the analyses were limited to Caucasians. HWE tests for each of the four SNPs were performed separately in the HPC probands, sporadic cases, and controls. The two SNPs in *HSD3B1* were in HWE in each of the groups. The two SNPs in *HSD3B2* (*B2-c7474t* and *B2-c7519g*) were in HWE in controls and in HPC probands but deviated from HWE in the sporadic cases ($P = 0.004$ and $P = 0.004$, respectively). Pair-wise LD tests for all four SNPs were also performed separately in each group. In all groups, the two SNPs within each gene were in strong LD ($P < 0.00001$), but the SNPs between the genes were either in weak LD ($P = 0.01$ between B1-N367T and *B2-c7474t* in HPC probands) or in linkage equilibrium ($P = 0.18$ and 0.88, between B1-N367T and *B2-c7474t* in sporadic cases and in unaffected controls, respectively).

To test the main hypothesis, that HSD3B genes are associated with prostate cancer risk, we compared the allele and genotype frequencies

Table 1 *Primers used to sequence HSD3B1 and HSD3B2 promoter and coding regions*

Gene	Amplified region	Forward primer	Reverse primer	Annealing temperature
HSD3B1				
	Promoter	TGACCGTTGATTGTCTCTGTT	GCAGGAGTAGCTGAAAGAAAATG	60
	Exon 1 & 2	AGAGCAATGAGTACATGGCCA	GGAGCAATGAGTATGTGGCAG	58
	Exon 3	CTTGTTCTTTCCGTAGAATG	CAGTCTTGAACTCTCCTATTC	60
	Exon 4–1st fragment	TGAGTCTGTTACAACCACCAT	AGTAGAACTGTCCTCGGATG	58
	Exon 4–2nd fragment	TTGTGCCTTACGACCCATG	CTTATAAGAGAAGGTGAATACG	60
	Exon 4–3rd fragment	CAGGCCAATTTACACCTATCG	TCAAACTATGTGAAGGAATGGA	60
HSD3B2				
	Promoter	GATTGGAGCTGTCACCATTG	CCTTACTGCCTCATCCCTG	60
	Exon 1 & 2	GGTCCATCTCCCCCCACATA	AGGTCAACCTCCCCACACC	68
	Exon 3	GGAATGTAGTACACCCTCCA	GCCTTGAACTCCCCAGTCA	60
	Exon 4–1st fragment	TGAGTCTGTTATAACCACTGC	TAATAGAATTGACCTCGGACA	60
	Exon 4–2nd fragment	CTTGTGCGTTAAGACCCACA	CTTGTAAGAGAAGGTGAACACA	60
	Exon 4–3rd fragment	CAGCCCAATTTACTCCTATCA	TTGAACTGTGTGAAGGAATAGG	60

Table 2 *Frequencies of HSD3B1 and HSD3B2 SNPs in 96 sequenced subjects*

Gene	SNP	Location of SNP	Nucleotide change	No. of chromosomes carrying variant in Caucasians: HPC probands (n = 48)	Sporadic cases (n = 48)	Controls (n = 48)	No. of chromosomes carrying variant in African-Americans: HPC probands (n = 16)	Sporadic cases (n = 16)	Controls (n = 16)
HSD3B1									
	B1-F286L	Exon 4	T to C	0	1	0	5	4	4
	B1-g6989a[a]	Exon 4	G to A	0	0	0	2	2	2
	B1-c7062t[a]	Exon 4	C to T	24	24	20	2	2	6
	B1-N367T	Exon 4	A to C	14	12	11	2	5	2
	B1-g7444a	3′-UTR	G to A	0	0	1	0	0	0
HSD3B2									
	B2-g(-322)a	Promoter	G to A	0	0	1	0	0	0
	B2-t4272a	Intron 3	T to A	1	1	0	0	0	0
	B2-c7294g	3′-UTR	C to G	0	1	1	0	0	3
	B2-a7400g	3′-UTR	A to G	0	1	1	0	0	3
	B2-c7474t	3′-UTR	C to T	8	5	5	3	2	10
	B2-c7519g	3′-UTR	C to G	7	5	5	2	1	5

[a] The nucleotide changes are synonymous with the SNPs.

for each of the four SNPs in HPC probands, sporadic cases, and unaffected controls (Table 3). Although variant alleles of three SNPs were observed at higher frequencies in cases than in controls, only one of them (the missense change, *B1-N367T*) reached nominal significance. The frequency of allele "*C* " of *B1-N367T* was higher in the HPC probands (34%) and in the sporadic cases (33%), compared with the unaffected controls (26%). The differences were significant between HPC probands and controls (P = 0.03), sporadic cases and controls (P = 0.04), and either type of prostate cancer and controls (P = 0.02). When the genotype frequencies of the four SNPs were compared, similar findings were observed (Table 4). The frequencies of the variant genotypes (*C/A* and *C/C*) of *B1-N367T* were higher in both HPC cases (55%) and sporadic cases (54%) than in the controls (43%). Compared with men with the wild-type genotype at *B1-N367T* (*A/A*), men with the variant genotypes at *B1-N367T* (*C/A* or *C/C*) were at increased risk for prostate cancer. After adjustment for age, the point estimate of the RR was 1.52 (95% CI = 0.95–2.45) for HPC, 1.5 (95% CI = 1.01–2.24) for sporadic prostate cancer, and 1.5 (95% CI = 1.04–2.17, P = 0.03) for either type of prostate cancer. In *HSD3B2*, the frequencies of the variant genotypes at *B2-c7474g* and *B2-c7519g* were also slightly higher in both the HPC cases and sporadic cases, compared with the controls, although the differences were not statistically significant.

We further tested the secondary hypothesis that the joint effect of the two genes is associated with prostate cancer risk. We were interested in testing two simple joint effects: whether men with a variant allele at: (*a*) either *HSD3B1* or *HSD3B2*; and (*b*) both *HSD3B1* and *HSD3B2* are at increased risk for prostate cancer compared with men who are homozygotes for the wild-type allele at both genes. However, because of the small number of people with variant genotypes at both genes, we did not have adequate power to test for the second joint effect. To test for the first joint effect, we calculated the proportion of men with the variant genotypes at either *B1-N367T* (*C/A or C/C*) or *B2-c7519g* (*C/G* or *G/G*) in HPC probands (74%), sporadic cases (68%), and unaffected controls (57%; Table 5). After adjustment for age, the differences were statistically significant between HPC probands and controls (P = 0.004), sporadic cases and controls (P = 0.02), and either type of prostate cancer and controls (P = 0.003). Compared with men with wild-type genotypes at both *B1-N367T* (*A/A*) and *B2-c7519g* (*C/C*), the age-adjusted point estimates of RR for HPC, sporadic prostate cancer, and either type of prostate cancer were 2.17 (95% CI = 1.29–3.65), 1.61 (95% CI = 1.07–2.42), and 1.76 (95% CI = 1.21–2.57), respectively, for men with the variant genotypes at either *B1-N367T* (*C/A* or *C/C*) or *B2-c7519g* (*C/G* or *G/G*).

The evidence for linkage at 1p13 is one of the two reasons that we were interested in the *HSD3B* genes. Because of this linkage evidence, and the stronger association between *HSD3B* genes and prostate cancer risk in our HPC probands, we explored whether the association was stronger in the subset of unrelated probands (n = 66) whose families provided evidence for linkage at 1p13 (LOD > 0). Results showed a stronger prostate cancer association with *B1-N367T*, and the joint effect of *B1-N367T* and *B2-c7519g*, in this subset sample. In this subset of 66 HPC probands, the remaining 67 probands, and unaffected controls, the proportion of men with the variant genotypes at *B1-N367T* was 61, 50, and 43%, respectively. The difference between the 66 HPC probands and controls was nominally significant (P = 0.04). The proportion of men with the variant genotypes at either *B1-N367T* (*C/A* or *C/C*) or B2-c7519g (*C/G* or *G/G*) was 78, 71, and 57% in the 66 HPC probands, the remaining 67 probands, and controls, respectively. The difference between the 66 HPC probands and controls was significant (P = 0.008). Thus, the subset of HPC probands whose families provided evidence for linkage at 1p13 predominantly contributed to the observed association. This also suggests that the evidence for linkage at 1p13 may be at least partially explained by the variants of *HSD3B* genes.

Considering that the younger controls may have a higher chance of developing prostate cancer later in their life than older controls because of the age-dependent penetrance of the disease, and that the evidence for linkage at 1p13 is provided primarily by families with

Table 3 *Allele frequencies of SNPs in HSD3B1 and HSD3B2 (Caucasians only)*

SNPs	Allele	HPC: No. of alleles	Sporadic: No. of alleles	Control: No. of alleles	χ^2 (P)[a]	χ^2 (P)[b]	χ^2 (P)[c]
HSD3B1-c7062t	T	97 (42%)	188 (44%)	160 (46%)	0.86 (.35)	0.14 (.70)	0.28 (.59)
HSD3B1-N367T	C	81 (34%)	147 (33%)	92 (26%)	4.94 (.03)	4.23 (.04)	5.76 (.02)
HSD3B2-c7474t	T	33 (14%)	59 (13%)	44 (12%)	0.67 (.41)	0.24 (.64)	0.42 (.52)
HSD3B2-c7519g	G	32 (14%)	62 (14%)	42 (11%)	0.82 (.37)	0.86 (.35)	1.01 (.31)

[a] HPC probands *vs.* controls.
[b] Sporadic cases *vs.* controls.
[c] Combined cases *vs.* controls.

Table 4 *Genotype frequencies of SNPs in HSD3B1 and HSD3B2 and association with prostate cancer (Caucasians only)*

SNPs	No. of subjects			RR[a] (95% CI)		
	HPC proband	Sporadic	Control	HPC *vs.* controls	Sporadic *vs.* controls	All cases *vs.* controls
HSD3B1-c7062t						
C/C	39 (34%)	65 (30%)	54 (31%)	1	1	1
C/T	57 (49%)	114 (53%)	82 (47%)			
T/T	20 (17%)	37 (17%)	39 (22%)	0.80 (0.39–1.60)	0.79 (0.44–1.41)	0.80 (0.47–1.34)
Any T	77 (66%)	151 (70%)	121 (69%)	0.94 (0.57–1.57)	1.04 (0.67–1.60)	1.01 (0.67–1.49)
HSD3B1-N367T						
A/A	53 (45%)	103 (46%)	99 (56%)	1	1	1
A/C	49 (41%)	97 (43%)	64 (36%)			
C/C	16 (14%)	25 (11%)	14 (8%)	2.09 (0.94–4.65)	1.71 (0.84–3.48)	1.81 (0.93–3.49)
Any C	65 (55%)	122 (54%)	78 (43%)	1.52 (0.95–2.45)	1.50 (1.01–2.24)	1.50 (1.04–2.17)
HSD3B2-c7474t						
C/C	86 (73%)	173 (78%)	146 (78%)	1	1	1
C/T	29 (25%)	41 (18%)	36 (20%)			
T/T	2 (2%)	9 (4%)	4 (2%)	1.03 (0.18–5.90)	1.90 (0.57–6.30)	1.56 (0.48–5.00)
Any T	31 (27%)	50 (22%)	40 (22%)	1.37 (0.79–2.38)	1.06 (0.66–1.69)	1.15 (0.74–1.76)
HSD3B2-c7519g						
C/C	85 (74%)	171 (77%)	146 (79%)	1	1	1
C/G	28 (24%)	42 (19%)	34 (19%)			
G/G	2 (2%)	10 (4%)	4 (2%)	0.94 (0.16–5.35)	2.14 (0.65–6.85)	1.70 (0.54–5.38)
Any G	30 (26%)	52 (23%)	38 (21%)	1.44 (0.82–2.52)	1.17 (0.72–1.84)	1.24 (0.80–1.92)

[a] All RRs were age adjusted.

older mean age of onset, we performed an analysis in subjects who were age ≥60 years (age of diagnosis for affected or age at examination for unaffected). Larger differences in the proportion of men with either variant genotype of the two SNPs were observed among HPC probands (76%), sporadic cases (74%), and unaffected controls (51%). After adjustment for age, the differences were statistically significant between HPC probands and controls ($P = 0.002$), sporadic cases and controls ($P = 0.005$), and all cases and controls ($P = 0.0005$).

Association between Characteristics of Prostate Cancer and the SNPs. The relationships between the four frequent polymorphisms in *HSD3B* genes and Gleason scores or pathological stages in sporadic prostate cancer cases were also examined. No statistically significant differences in the genotypic frequencies of these SNPs were found between the groups with low (≤6) or high (≥7) Gleason scores or between the groups with disease confined to the prostate *versus* nonlocalized disease (Table 6).

DISCUSSION

HSD3B1 and *HSD3B2* are important candidate genes for prostate cancer susceptibility because of their biological importance in the androgen metabolism pathway and their chromosomal location. To assess the potential role of these genes in prostate cancer susceptibility, we screened DNA samples from men with or without prostate cancer for sequence variants in both genes. We identified a total of 11 SNPs, 4 of which are informative, for further analysis. Although a weak association between prostate cancer risk and a missense change in *HSD3B1* (*B1-N367T*) was suggested when each of the 4 informative SNPs were analyzed independently, stronger evidence for association was found when the joint effect of the two *HSD3B* genes were considered. Men with the variant genotypes at either *B1-N367T* or *B2-c7519g* had a RR of 1.76 (95% CI = 1.21–2.57, $P = 0.003$) for prostate cancer, compared with men who were homozygous wild type at both genes. The risk for HPC was stronger, with a RR of 2.17 (95% CI = 1.29–3.65, $P = 0.003$). Most importantly, the subset of HPC probands whose families provided evidence for linkage predominantly contributed to the observed association.

Although these results are potentially important, caution should be taken when interpreting and generalizing these findings. Our case-control population has several potential limitations. First of all, the study subjects were recruited primarily for genetics studies rather than for a rigorously designed epidemiological study. Thus, it is difficult to interpret the point estimates of the RR in this study and to generalize these findings. However, this study does provide some valuable results. The SNPs identified in our studies can be used in the future to study prostate cancer and other diseases. The increased frequencies of the variant *HSD3B* SNPs in the cases (particularly HPC cases) should prompt additional studies. The second potential limitation is the source of our control subjects, which were recruited from a prostate cancer screening population. This control group may represent a higher risk population than the general population because of self-selection. This potential bias, however, is unlikely to be significant in our study. All control subjects were found to have normal digital rectal examination and PSA results at the time of screening. Three percent of the 182 personally interviewed controls reported a positive family history (defined as an affected father and/or brothers). Additional analyses excluding the individuals who reported positive family history produced similar results. The third potential limitation in our study is that the association is subject to potential population stratification. Differences in the allele frequencies between cases and

Table 5 *Frequencies and RRs for combined HSD3B1-N367T and HSD3B2-c7474t genotypes (Caucasians only)*

SNPs	No. of subjects			RR[a] (95% CI)		
	HPC	Sporadic	Control	HPC *vs.* controls	Sporadic *vs.* controls	All cases *vs.* controls
All ages						
B1-N367T = A/A and B2-c7519g = C/C	30 (26%)	72 (32%)	77 (43%)	1	1	1
B1-N367T = A/C or C/C or B2-c7519g = C/G or G/G	84 (74%)	152 (68%)	101 (57%)	2.17 (1.29–3.65)	1.61 (1.07–2.42)	1.76 (1.21–2.57)
Older age group (>60 years)						
B1-N367T = A/A and B2-c7519g = C/C	17 (24%)	26 (26%)	33 (49%)	1	1	1
B1-N367T = A/C or C/C or B2-c7519g = C/G or G/G	55 (76%)	73 (74%)	34 (51%)	3.14 (1.52–6.49)	2.62 (1.34–5.14)	2.88 (1.59–5.23)

[a] All odds ratios were age adjusted.

Table 6 *Genotype frequencies of SNPs in HSD3B1 and HSD3B2 in sporadic cases (Caucasians only)*

	Glenson score, *n* (%)		Pathological stage, *n* (%)	
	≤6	≥7	0	≥1
HSD3B1-c7062t				
C/C	28 (34.14)	37 (27.61)	26 (38.23)	39 (26.35)
C/T	42 (51.22)	72 (53.73)	34 (50.00)	80 (54.05)
T/T	12 (14.63)	25 (18.66)	8 (11.76)	29 (19.59)
HSD3B1-N367T				
A/A	34 (42.50)	65 (49.62)	33 (49.25)	66 (45.83)
A/C	39 (48.75)	52 (39.69)	28 (41.79)	63 (43.75)
C/C	7 (8.75)	14 (10.69)	6 (8.95)	15 (10.42)
HSD3B2-c7474t				
C/C	65 (77.38)	108 (77.70)	53 (79.10)	120 (76.92)
C/T	16 (19.05)	25 (17.98)	11 (16.42)	30 (19.23)
T/T	3 (3.57)	6 (4.32)	3 (4.48)	6 (3.85)
HSD3B2-c7519g				
C/C	61 (74.39)	110 (78.01)	52 (78.79)	119 (75.80)
C/G	18 (21.95)	24 (17.02)	11 (16.67)	31 (19.74)
G/G	3 (3.66)	7 (4.96)	3 (4.54)	7 (4.46)

controls could be attributable to the different genetic backgrounds in cases and controls. We attempted to limit the impact of this source of population stratification by limiting our analyses to Caucasian men only, although this approach might not fully remove the potential impact. On the other hand, based on a sample of 24 consecutive SNPs on chromosomes 1, 8, 11, 12, and X that were recently genotyped in this population, we found no evidence to suggest population stratification exists within our Caucasian case and control samples (data not shown). A family-based association test is an alternative study design to overcome the potential bias of population stratification. However, a family-based association study is inefficient in this population because most parents of affected men are deceased because of the late age of onset of prostate cancer. The fourth potential limitation is the multiple tests performed in our study. Not only were multiple SNPs genotyped, but multiple hypotheses (dominant or recessive and single SNP or joint effect) and multiple groups (HPC probands, sporadic cases, and unaffected controls) were also tested for each of the SNPs. Some of the tests are not independent, and appropriate methods are not available to adjust the significance level because of the multiple but related comparisons. However, using the commonly suggested Bonferoni test, we calculated adjusted significance levels by multiplying the nominal *P*s by the total number of tests performed in the study ($n = 44$). After the adjustment, the only statistically significant finding was the association between prostate cancer risk and the joint effect of the two genes. With these caveats, we cautiously report our findings and call for large well-designed studies to rigorously evaluate these findings.

The hypothesis that sequence variants in either HSD3B1 or HSD3B2 may increase prostate cancer susceptibility is biologically plausible; however, the exact mechanism by which such an effect may be mediated is not defined. HSD3B genes encode membrane-bound microsomal proteins with two predicted transmembrane domains: (*a*) a 16-residue segment between residues 75 and 91; and (*b*) a COOH-terminal 26-residue segment between residues 283 and 308. The *B1-N367T* variant is located in the COOH-terminal extramembrane domain. This SNP results in an amino acid change from Asn to Thr and may have an effect on conformation, enzymatic activity, stability, or regulation of HSD3B1 protein. This amino acid change creates a new putative PKC phosphorylation site (the phosphorylation site pattern: [ST][.][RK]).[4] PKC isozymes are a family of kinases in the signal transduction cascade and are involved in cell proliferation, antitumor resistance, and apoptosis. It has been shown that *HSD3B1* gene expression is specifically induced by IL-4 and IL-13 in both human prostate cancer cell lines and primary prostatic epithelial cells (21). In addition, the PKC activator phorbol-12-myristate-13-acetate further enhanced the stimulatory effect of *IL-4* on *HSD3B* activity (34). It is possible that HSD3B proteins are regulated through phosphorylation by PKC, and it is worth exploring whether the new PKC phosphorylation site in a variant HSD3B1 protein alters the regulation of HSD3B1 protein. Because SNP *B2-c7519g* is located in the 3′-UTR of *HSD3B2,* it has no effect on the amino acid sequence of HSD3B2 protein. However, the nucleotide change may result in a conformational change in the 3′-UTR of *HSD3B2* mRNA and may affect the stability of this mRNA. Post-transcriptional regulation of mRNA stability can have a significant impact on mRNA abundance and subsequent protein expression. Several elements in the 3′-UTR region that are important to the stability of a variety of mRNA species have been identified, including the poly(A) site, arbitrary unit-rich elements, iron-responsive element, 3′-terminal stem-loop, long-range stem loop, exoribonuclease cleavage site, and endoribonuclease cleavage site. It is possible that the nucleotide change in the 3′-UTR of *HSD3B2* mRNA alters the structure of a protein binding site and, hence, alters the stability of the mRNA and the quantity of the protein produced.

We tested the secondary hypothesis that the joint effect of the two genes is associated with prostate cancer risk for the following two reasons: (*a*) even with the similarity in the structure and enzymatic function between HSD3B1 and HSD3B2 proteins, the differential expression patterns of *HSD3B1* and *HSD3B2* genes in different tissues implicate HSD3B1 and HSD3B2 as being involved in the regulation of androgen levels in different ways. HSD3B2, which is predominately expressed in steroidogenic tissues, may be more important for systematic androgen levels. On the other hand, HSD3B1, which is primarily expressed in peripheral tissues, including prostate, may play a more important role in local androgen levels; and (*b*) if either variant at HSD3B1 or HSD3B2 increases the risk for prostate cancer, a single SNP analysis would be a less powerful approach when the two genes are not in complete LD. This is because the genotypes at the other gene (SNP) may confound the effect of the genotypes at the gene (SNP) under study. This confounding effect can be decreased by studying the two genes (SNPs) simultaneously. Whereas the false positive rate is not increased when there is no association between a disease and either gene, these analyses do increase the total number of tests and, thus, affect the interpretation of significance level.

Consistent with the results of our previous linkage study, where families with late age of diagnosis of prostate cancer have the strongest evidence for linkage to the region of *HSD3B* genes (16), the highest risk (odds ratio = 3.14) for HPC was observed in the men with late age of onset in the present study. Although the reason for this finding is unknown, genetic heterogeneity could partially explain this observation. Several other prostate cancer susceptibility genes have been reported, including *HPC1* at 1q24–25 (9), *PCAP* at 1q42–43 (10), *HPCX* at Xq27–28 (11), *CAPB* at 1p36 (12), *HPC20* at 20q13 (13), and *HPC2/ELAC2* on chromosome 17 (14). Evidence for linkage to some of these regions has primarily been observed in prostate cancer families with early age of onset, *e.g.,* the linkage study of chromosome 1 markers in our 159 HPC families only observed linkage at *HPC1* in the 79 families with early age of onset, with a peak allele sharing LOD of 3.05 ($P = 0.0002$). However, the 80 families with late age of onset were not linked to *HPC1*.

The deviation from HWE for the two SNPs of *HSD3B2* (*B2-c7474t* and *B2-c7519g*) in sporadic prostate cancer cases is an interesting result. This result is unlikely attributable to genotyping errors, because the SNPs were unambiguously scored by three experienced molecular geneticists (B-l. C., G. A. H., and S. L. Z), and the distributions of the two closely linked SNPs were very similar (Table 4). Two other

[4] Internet address: http://maple.bioc.columbia.edu/predictprotein/.

explanations are possible: (*a*) the deviation from HWE could be attributable to chance; the observed number of homozygotes of the rare alleles (9 and 10 in *B2-c7474t* and B2-c7519g, respectively) is only slightly more than the expected number of 6; and (*b*) the two sequence variants may be either causal changes or in strong LD with a causal change.

In summary, our study provides evidence for association between HSD3B genes and prostate cancer risk. Considering the importance of this gene family, the complexities of the genetics of prostate cancer, and the limitations of our study, additional studies at a functional level, as well as additional study populations, are warranted.

ACKNOWLEDGMENTS

We thank all of the subjects who participated in this study.

REFERENCES

1. Isaacs, W. B., Xu, J., and Walsh, P. C., Hereditary prostate cancer. *In:* L. W. K. Chung, W. B. Isaacs, and J. W. Simons (eds.), Prostate Cancer, Biology, Genetics, and the New Therapeutics. Totowa, NJ: Humana Press, 2001.
2. Gronberg, H., Damber, L., and Damber, J. E. Studies of genetic factors in prostate cancer in a twin population. J. Urol., *152:* 1484–1487, 1994.
3. Page, W. F., Braun, M. M., Partin, A. W., Caporaso, N., and Walsh, P. Heredity and prostate cancer: a study of World War II veteran twins. Prostate, *33:* 240–245, 1997.
4. Lichtenstein, P., Holm, N. V., Verkasalo, P. K., Iliadou, A., Kaprio, J., Koskenvuo, M., Pukkala, E., Skytthe, A., and Hemminki, K. Environmental and heritable factors in the causation of cancer–analyses of cohorts of twins from Sweden, Denmark, and Finland. N. Engl. J. Med., *343:* 78–85, 2000.
5. Carter, B. S., Beaty, T. H., Steinberg, G. D., Childs, B., and Walsh, P. C. Mendelian inheritance of familial prostate cancer. Proc. Natl. Acad. Sci. USA, *89:* 3367–3371, 1992.
6. Gronberg, H., Damber, L., Damber, J. E., and Iselius, L. Segregation analysis of prostate cancer in Sweden: support for dominant inheritance. Am. J. Epidemiol., *146:* 552–557, 1997.
7. Schaid, D. J., McDonnell, S. K., Blute, M. L., and Thibodeau, S. N. Evidence for autosomal dominant inheritance of prostate cancer. Am. J. Hum. Genet., *62:* 1425–1438, 1998.
8. Cui, J., Staples, M. P., Hopper, J. L., English, D. R., McCredie, M. R., and Giles, G. G. Segregation analyses of 1,476 population-based Australian families affected by prostate cancer. Am. J. Hum. Genet., *68:* 1207–1218, 2001.
9. Smith, J. R., Freije, D., Carpten, J. D., Gronberg, H., Xu, J., Isaacs, S. D., Brownstein, M. J., Bova, G. S., Guo, H., Bujnovszky, P., Nusskern, D. R., Damber, J. E., Bergh, A., Emanuelsson, M., Kallioniemi, O. P., Walker-Daniels, J., Bailey-Wilson, J. E., Beaty, T. H., Meyers, D. A., Walsh, P. C., Collins, F. S., Trent, J. M., and Isaacs, W. B. Major susceptibility locus for prostate cancer on chromosome 1 suggested by a genome-wide search. Science (Wash. DC), *274:* 1371–1374, 1996.
10. Berthon, P., Valeri, A., Cohen-Akenine, A., Drelon, E., Paiss, T., Wohr, G., Latil, A., Millasseau, P., Mellah, I., Cohen, N., Blanche, H., Bellane-Chantelot, C., Demenais, F., Teillac, P., Le Duc, A., de Petriconi, R., Hautmann, R., Chumakov, I., Bachner, L., Maitland, N. J., Lidereau, R., Vogel, W., Fournier, G., Mangin, P., Cohen, D., and Cussenot, O. Predisposing gene for early-onset prostate cancer, localized on chromosome 1q42.2–43. Am. J. Hum. Genet., *62:* 1416–1424, 1998.
11. Xu, J., Meyers, D., Freije, D., Isaacs, S., Wiley, K., Nusskern, D., Ewing, C., Wilkens, E., Bujnovszky, P., Bova, G. S., Walsh, P., Isaacs, W., Schleutker, J., Matikainen, M., Tammela, T., Visakorpi, T., Kallioniemi, O. P., Berry, R., Schaid, D., French, A., McDonnell, S., Schroeder, J., Blute, M., Thibodeau, S., Gronberg, H., Emanuelsson, M., Damber, J., Bergh, A., Jonsson, B., Smith, J., Bailey-Wilson, J., Carpten, J., Stephan, D., Gillanders, E., Amundson, I., Kainu, T., Freas-Lutz, D., Baffoe-Bonnie, A., Aucken, A. V., Sood, R., Collin, F., Brounstein, M., and Trent, J. Evidence for a prostate cancer susceptibility locus on the X chromosome. Nat. Genet., *20:* 175–179, 1998.
12. Gibbs, M., Stanford, J. L., McIndoe, R. A., Jarvik, G. P., Kolb, S., Goode, E. L., Chakrabarti, L., Schuster, E. F., Buckley, V. A., Miller, E. L., Brandzel, S., Li, S., Hood, L., and Ostrander, E. A. Evidence for a rare prostate cancer-susceptibility locus at chromosome 1p36. Am. J. Hum. Genet., *64:* 776–787, 1999.
13. Berry, R., Schroeder, J. J., French, A. J., McDonnell, S. K., Peterson, B. J., Cunningham, J. M., Thibodeau, S. N., and Schaid, D. J. Evidence for a prostate cancer-susceptibility locus on chromosome 20. Am. J. Hum. Genet., *67:* 82–91, 2000.
14. Tavtigian, S. V., Simard, J., Teng, D. H., Abtin, V., Baumgard, M., Beck, A., Camp, N. J., Carillo, A. R., Chen, Y., Dayananth, P., Desrochers, M., Dumont, M., Farnham, J. M., Frank, D., Frye, C., Ghaffari, S., Gupte, J. S., Hu, R., Iliev, D., Janecki, T., Kort, E. N., Laity, K. E., Leavitt, A., Leblanc, G., McArthur-Morrison, J., Pederson, A., Penn, B., Peterson, K. T., Reid, J. E., Richards, S., Schroeder, M., Smith, R., Snyder, S. C., Swedlund, B., Swensen, J., Thomas, A., Tranchant, M., Woodland, A. M., Labrie, F., Skolnick, M. H., Neuhausen, S., Rommens, J., and Cannon-Albright, L. A. A candidate prostate cancer susceptibility gene at chromosome 17p. Nat. Genet., *27:* 172–180, 2001.
15. Rebbeck, T. R., Walker, A. H., Zeigler-Johnson, C., Weisburg, S., Martin, A. M., Nathanson, K. L., Wein, A. J., and Malkowicz, S. B. Association of HPC2/ELAC2 genotypes and prostate cancer. Am. J. Hum. Genet., *67:* 1014–1019, 2001.
16. Xu, J., Zheng, S. L., Hawkins, G. A., Faith, D. A., Kelly, B., Isaacs, S. D., Wiley, K. E., Chang, B., Ewing, C. M., Bujnovszky, P., Carpten, J. D., Bleecker, E. R., Walsh, P. C., Trent, J. M., Meyers, D. A., and Isaacs, W. B. Linkage and association studies of prostate cancer susceptibility: evidence for linkage at 8p22–23. Am. J. Hum. Genet., *69:* 341–350, 2001.
17. Luu, T., V, Lachance, Y., Labrie, C., Leblanc, G., Thomas, J. L., Strickler, R. C., and Labrie, F. Full length cDNA structure and deduced amino acid sequence of human 3 β-hydroxy-5-ene steroid dehydrogenase. Mol. Endocrinol., *3:* 1310–1312, 1989.
18. Rheaume, E., Lachance, Y., Zhao, H. F., Breton, N., Dumont, M., de Launoit, Y., Trudel, C., Luu-The, V., Simard, J., and Labrie, F. Structure and expression of a new complementary DNA encoding the almost exclusive 3 β-hydroxysteroid dehydrogenase/Δ 5-Δ 4-isomerase in human adrenals and gonads. Mol. Endocrinol., *5:* 1147–1157, 1991.
19. McBride, M. W., McVie, A. J., Burridge, S. M., Brintnell, B., Craig, N., Wallace, A. M., Wilson, R. H., Varley, J., and Sutcliffe, R. G. Cloning, expression, and physical mapping of the 3β-hydroxysteroid dehydrogenase gene cluster (HSD3BP1-HSD3BP5) in human. Genomics, *61:* 277–284, 1999.
20. Simard, J., Durocher, F., Mebarki, F., Turgeon, C., Sanchez, R., Labrie, Y., Couet, J., Trudel, C., Rheaume, E., Morel, Y., Luu-The, V., and Labrie, F. Molecular biology and genetics of the 3 β-hydroxysteroid dehydrogenase/Δ5-Δ4 isomerase gene family. J. Endocrinol., *150* (Suppl.): S189–S207, 1996.
21. Gingras, S., and Simard, J. Induction of 3β-hydroxysteroid dehydrogenase/isomerase type 1 expression by interleukin-4 in human normal prostate epithelial cells, immortalized keratinocytes, colon, and cervix cancer cell lines. Endocrinology, *140:* 4573–4584, 1999.
22. Gingras, S., Moriggl, R., Groner, B., and Simard, J. Induction of 3β-hydroxysteroid dehydrogenase/Δ5-Δ4 isomerase type 1 gene transcription in human breast cancer cell lines and in normal mammary epithelial cells by interleukin-4 and interleukin-13. Mol. Endocrinol., *13:* 66–81, 1999.
23. El Alfy, M., Luu-The, V., Huang, X. F., Berger, L., Labrie, F., and Pelletier, G. Localization of type 5 17β-hydroxysteroid dehydrogenase, 3β-hydroxysteroid dehydrogenase, and androgen receptor in the human prostate by *in situ* hybridization and immunocytochemistry. Endocrinology, *140:* 1481–1491, 1999.
24. Rheaume, E., Simard, J., Morel, Y., Mebarki, F., Zachmann, M., Forest, M. G., New, M. I., and Labrie, F. Congenital adrenal hyperplasia due to point mutations in the type II 3 β-hydroxysteroid dehydrogenase gene. Nat. Genet., *1:* 239–245, 1992.
25. Xu, J., Zheng, S. L., Chang, B., Smith, J. R., Carpten, J. D., Stine, O. C., Isaacs, S. D., Wiley, K. E., Henning, L., Ewing, C., Bujnovszky, P., Bleeker, E. R., Walsh, P. C., Trent, J. M., Meyers, D. A., and Isaacs, W. B. Linkage of prostate cancer susceptibility loci to chromosome 1. Hum. Genet., *108:* 335–345, 2001.
26. Goddard, K. A., Witte, J. S., Suarez, B. K., Catalona, W. J., and Olson, J. M. Model-free linkage analysis with covariates confirms linkage of prostate cancer to chromosomes 1 and 4. Am. J. Hum. Genet., *68:* 1197–1206, 2001.
27. Verreault, H., Dufort, I., Simard, J., Labrie, F., and Luu-The, V. Dinucleotide repeat polymorphisms in the *HSD3B2* gene. Hum. Mol. Genet., *3:* 384, 1994.
28. Devgan, S. A., Henderson, B. E., Yu, M. C., Shi, C. Y., Pike, M. C., Ross, R. K., and Reichardt, J. K. Genetic variation of 3 β-hydroxysteroid dehydrogenase type II in three racial/ethnic groups: implications for prostate cancer risk. Prostate, *33:* 9–12, 1997.
29. Lachance, Y., Luu-The, V., Verreault, H., Dumont, M., Rheaume, E., Leblanc, G., and Labrie, F. Structure of the human type II 3 β-hydroxysteroid dehydrogenase/Δ 5-Δ 4 isomerase (3 β-HSD) gene: adrenal and gonadal specificity. DNA Cell Biol., *10:* 701–711, 1991.
30. Lachance, Y., Luu-The, V., Labrie, C., Simard, J., Dumont, M., de Launoit, Y., Guerin, S., Leblanc, G., and Labrie, F. Characterization of human 3 β-hydroxysteroid dehydrogenase/Δ 5- Δ 4-isomerase gene and its expression in mammalian cells. J. Biol. Chem., *267:* 3551, 1992.
31. Weir, B. S. (ed.). Genetic Data Analysis II: Methods for Discrete Population Genetic Data. Sunderland, MA: Sinauer Association, Inc. Publishers, 1996.
32. Zaykin, D., Zhivotovsky, L., and Weir, B. S. Exact tests for association between alleles at arbitrary numbers of loci. Genetica, *96:* 169–178, 1995.
33. Schlesselman, J. J. (ed.). Case-control Studies. New York: Oxford University Press, 1982.
34. Gingras, S., Cote, S., and Simard, J. Multiple signal transduction pathways mediate interleukin-4-induced 3β-hydroxysteroid dehydrogenase/Δ5-Δ4 isomerase in normal and tumoral target tissues. J. Steroid Biochem. Mol. Biol., *76:* 213–225, 2001.

Hum Genet (2002) 110:122–129
DOI 10.1007/s00439-001-0662-6

ORIGINAL INVESTIGATION

Bao-li Chang · Siqun L. Zheng · Gregory A. Hawkins · Sarah D. Isaacs · Kathy E. Wiley · Aubrey Turner John D. Carpten · Eugene R. Bleecker · Patrick C. Walsh · Jeffrey M. Trent · Deborah A. Meyers William B. Isaacs · Jianfeng Xu

Polymorphic GGC repeats in the androgen receptor gene are associated with hereditary and sporadic prostate cancer risk

Received: 6 September 2001 / Accepted: 13 November 2001 / Published online: 23 January 2002

Abstract Androgen receptor (AR) has long been hypothesized to play an important role in prostate cancer etiology. Two trinucleotide repeat polymorphisms (CAG and GGC repeats in exon 1 of the AR gene) have been investigated as risk factors for prostate cancer in several studies. However, the results are inconclusive, probably because of the variations of study designs, characteristics of study samples, and choices of analytical methods. In this study, we evaluated evidence for linkage and association between the two AR repeats and prostate cancer by using the following comprehensive approaches: (1) a combination of linkage and association studies, (2) a test for linkage by parametric analysis and the male-limited X-linked transmission/disequilibrium test (XLRC-TDT), (3) a test for association by using both population-based and family-based tests, and (4) a study of both hereditary and sporadic cases. A positive but weak linkage score (HLOD=0.49, P=0.12) was identified in the AR region by parametric analysis; however, stronger evidence for linkage in the region, especially at the GGC locus, was observed in the subset of families whose proband had ≤16 GGC repeats (HLOD=0.70, P=0.07) or by using XLRC-TDT (z'=2.65, P=0.008). Significantly increased frequencies of the ≤16 GGC repeat alleles in 159 independent hereditary cases (71%) and 245 sporadic cases (68%) cases compared with 211 controls (59%) suggested that GGC repeats were associated with prostate cancer (P=0.02). Evidence for the association between the ≤16 GGC repeats and prostate cancer risk was stronger with XLRC-TDT (z'=2.66, P=0.007). No evidence for association between the CAG repeats and prostate cancer risk was observed. The consistent results from both linkage and association studies strongly implicate the GGC repeats in the AR as a prostate cancer susceptibility gene. Further studies on this polymorphism in other independent data sets and functional analysis of the GGC repeat length on AR activity are warranted.

B.-l. Chang · S.L. Zheng · G.A. Hawkins · A. Turner
E.R. Bleecker · D.A. Meyers · J. Xu (✉)
Center for Human Genomics,
Wake Forest University School of Medicine,
Medical Center Boulevard, Winston-Salem, NC 27157, USA
e-mail: jxu@wfubmc.edu,
Tel.: +1-336-7165700, Fax: +1-336-7167575

B.-l. Chang
University of Maryland School of Medicine,
Baltimore, Md., USA

S.D. Isaacs · K.E. Wiley · P.C. Walsh · W.B. Isaacs
Department of Urology, Johns Hopkins Medical Institutions,
Baltimore, Md., USA

J.D. Carpten · J.M. Trent
National Human Genome Research Institute,
National Institute of Health, Bethesda, Md., USA

Introduction

Although genetic susceptibility to prostate cancer has been well established, the modes of inheritance of prostate cancer appear complex. Whereas an autosomal dominant mode of inheritance has been suggested from several segregation studies (Carter et al. 1992; Gronberg et al. 1997; Schaid et al. 1998; Cui et al. 2001), an X-linked or recessive mode of inheritance has also been implicated (Cui et al. 2001). Several population-based studies have also reported a statistically significant excess risk of prostate cancer in men with affected brothers compared with those with affected fathers, consistent with the hypothesis of an X-linked or recessive mode of inheritance (Woolf 1960; Narod et al. 1995; Hayes et al. 1995; Monroe et al. 1995; Cerhan et al. 1999; Schuurman et al. 1999). Results from linkage studies have provided further evidence of a prostate cancer susceptibility locus, HPCX, at the q27–28 region of the X-chromosome (Xu et al. 1998).

Androgens have long been hypothesized to be involved in prostate carcinogenesis because of their essential role in prostate development, growth, and maintenance. The androgen receptor (AR) gene, located on Xq11–12 (~50 cM centromeric to HPCX), is a compelling candidate gene for prostate cancer. The AR gene encodes for a transcription factor within the steroid receptor superfamily (Chang et al. 1988; Lubahn et al. 1988). To date, more than 50 somatic

mutations of AR have been found in clinical specimens of prostate cancer (http://www.mcgill.ca/androgendb). In addition, germ-line mutations of the AR gene have also been reported (Elo et al. 1995; Crocitto et al. 1997; Mononen et al. 2000).

One critical function of the AR gene product is to activate the expression of other genes. The transactivation activity resides in the N-terminal domain of the protein, encoded by exon 1. Two polymorphic microsatellites are located approximately 1.1 kb apart in exon 1: a highly polymorphic CAG repeat and a less polymorphic GGC repeat (Edwards et al. 1992; Sleddens et al. 1993). The CAG repeat encodes a poly-glutamine tract and usually contains 9–29 repeats (Irvine et al. 1995). Alleles of the GGC repeat code for a polyglycine tract and contain 4–21 repeats, with 16 repeats being the most common allele. Coetzee and Ross (1994) have suggested that enhanced activity of the AR, attributable to polymorphisms in the AR gene, might alter the risk of prostate cancer. An inverse correlation between the length of CAG repeat and the transactivation activities of AR has been demonstrated by several in vitro assays (Mhatre et al. 1993; Chamberlain et al. 1994; Beilin et al. 2000). However, there is no report of an association between the length of GGC repeats and functional changes of AR.

Several linkage studies have tested the hypothesis that AR is a prostate susceptibility gene. Results from two large-scale prostate cancer genome-wide screens have provided evidence for linkage at the AR region. In a study by Goddard et al. (2001), a LOD of 3.06 (P=0.0005) at the AR region was reported in 254 families after the Gleason score was included as a covariate. In the study by Hsieh et al. (2001), a multipoint non-parametric linkage (NPL) Z-score of 1.5 was observed in 98 multiple affected families. However, results from two other prostate cancer genome-wide screens failed to provide evidence for linkage at the AR region (66 families in Smith et al. 1996; and 94 families in Gibbs et al. 2000). In a study with the CAG repeats as a marker, no evidence for prostate cancer linkage was observed (Lange et al. 2000). The power to detect linkage of complex diseases with substantial locus heterogeneity is largely dependent on the sample size, informativeness of the families, resolution of markers, choice of study designs, and choice of analytical methods. Additional linkage studies with multiple markers at the AR in large and well-characterized prostate cancer families are warranted. Furthermore, because both CAG and GGC repeats could be associated with prostate cancer risk, tests for linkage in the presence of association by using the transmission/disequilibrium tests may have better power (Spielman et al. 1993; Knapp 1999).

The hypothesis that shorter alleles of CAG and (or) GGC repeats in the AR are associated with an increased risk for prostate cancer has also been tested in several association studies (Table 1). Three inferences can be draw from these studies. First, the association between the AR repeats and prostate cancer is inconclusive. Whereas some studies reported a marginally increased risk for individuals with short CAG repeats and/or short GGC repeats, an almost equal number of studies did not find a significant association. Second, among the studies that tested both CAG and GGC repeats, the association with prostate cancer risk was stronger with GGC repeats or combinations of GGC and CAG repeats (Hakimi et al. 1997; Stanford et al. 1997; Platz et al. 1998). For example, in a study of 301 prostate cancer cases and 277 controls, Stanford and colleagues (1997) did not find a significantly increased risk for men with ≤21 CAG repeats compared with men with ≥22 CAG repeats [relative risk (RR)=1.23, 95% confidence interval (CI)=0.88–1.73], but found a significantly increased risk for men with ≤16 GGC repeats compared with men who had ≥17 GGC repeats (RR=1.60, 95% CI=1.07–2.41). Third, few studies have evaluated and compared the risk of CAG repeats in hereditary prostate cancer (Lange et al. 2000), and no study has evaluated GGC repeats in hereditary prostate cancer. It is unclear whether the AR repeats impose a higher or lower cancer risk in hereditary prostate cancer compared with the sporadic form.

Considering the biological importance of AR in prostate cancer, the inconclusive results from the linkage and association studies of AR, and especially the fact that few studies have investigated and compared the risk of AR repeats in hereditary and sporadic prostate cancer, we have performed the following three analyses. First, we have tested for linkage between a prostate cancer susceptibility gene and AR repeats in 159 HPC families, each with at least three first-degree relatives being affected with prostate cancer. Second, we have tested for an association between AR repeats and either hereditary or sporadic prostate cancers by comparing the frequency of CAG and GGC repeats in the 159 HPC probands, 245 sporadic prostate cancer cases, and 211 unaffected controls. Last, we have performed a family-based linkage and association analysis to improve the power of linkage detection in the presence of association and to eliminate the potential confounder of population stratification.

Materials and methods

Nomenclature

Gene mutation nomenclature used in this article follows the recommendations of den Dunnen and Antonarakis (2001).

Subjects

All individuals in this study gave full informed consent. A detailed description of the study sample has been presented elsewhere (Xu et al. 2001). Briefly, a total of 159 HPC families were collected and studied at the Brady Urology Institute of Johns Hopkins Hospital (Baltimore, Md.). The diagnosis of prostate cancer was verified by medical records for each affected male studied. The mean age at diagnosis was 64.3 years; 84% of the families were Caucasian, 6.9% were Ashkenazi Jewish, and 8.8% were African American. The average number of affected individuals per family was 5.08. The number of families with 3, 4, and 5 or more affected individuals was 29, 40, and 90, respectively.

All 245 unrelated prostate cancer cases were recruited from patients who underwent treatment for prostate cancer at the Johns

Table 1 Reported association studies between prostate cancer risk and AR CAG and/or GGC repeats (*RR* relative risk, *CI* confidence interval)

Study population	No. subjects	RR (95% CI or *P*-value)	References
Non-Hispanic whites	57 sporadic cases 39 controls	CAG<22 and GGC-non-16 vs. others: 2.1 (*P*=0.08)	Irvine et al. 1995
US Caucasians	301 sporadic cases 277 controls	For CAG<22 vs CAG>=22:1.23 (0.88–1.73) For GGC<=16 vs GGC>16:1.6 (1.07–2.41) For CAG<22 and GGC<=16:2.05 (1.09–3.84)	Stanford et al. 1997
Predominantly Caucasians	587 sporadic cases 588 controls	CAG<=18 vs CAG>=26: 2.14 (1.14–4.01)	Giovannucci et al. 1997
Caucasians	59 sporadic cases 370 men from general population	CAG<=17 vs CAG>17:3.7 (1.3–10.5) GGC<=14 vs >CCG>14:4.6 (1.3–16.1)	Hakimi et al. 1997
Non-Hispanic whites	57 sporadic cases 169 controls	CAG<20 vs CAG>=20:2.10 (1.11–3.99)	Ingles et al. 1997
US Sweden	160 sporadic cases 186 controls	No significant association when CAG repeat length was categorized in tertiles (15–20; 21–23; 24–31)	Bratt et al. 1999
French-German	132 sporadic cases 105 controls	CAG<22 vs CAG>=22 1.2 (0.7–2.0)	Correa-Cerro et al. 1999
Caucasian	178 sporadic cases 195 controls	For CAG<=21 vs CAG>21:1.00 (0.96–1.03) For GGC<=16 vs GGC>16:1.06 (0.70–1.76)	Edwards et al. 1999
Chinese	190 sporadic cases 304 controls	For CAG<23 vs CAG>=23:1.65 (1.14–2.39) For GGC<=16 vs GGC>16:1.12 (0.71–1.78)	Hsing et al. 2000
Caucasian	133 cases with family history of prostate cancer 305 controls	For CAG<=21 vs CAG>21:0.90 (0.60–1.36) CAG<=18 vs CAG>=26:0.73 (0.31–1.69)	Lange et al. 2000
Caucasian	140 cases in 51 high-risk sibships 70 unaffected brothers of cases	For CAG<=21 vs CAG>21:1.13 (0.5–2.4) CAG<=16 vs CAG>16:0.98 (0.46–2.06)	Miller et al. 2001

Hopkins Hospital and did not have first-degree relatives affected with prostate cancer. For each subject, the diagnosis of prostate cancer was confirmed by pathology reports. Preoperative prostate specific antigen (PSA) levels, Gleason score, and pathological stages were available for 202, 240, and 241 cases, respectively. The mean age at diagnosis for these cases was 58.7 years. Over 93% of the cases were Caucasian, and 3.2% were African American.

Non-prostate cancer controls (n=222) were selected from men participating in screening programs for prostate cancer. By applying the exclusion criteria of abnormal digital rectal examination (DRE) and abnormal PSA level (i.e., >=4 ng/ml), 211 men were eligible for the study. The mean age at examination was 58 years. Over 86% of the eligible controls were Caucasian, and 7.1% were African American. In interviews, approximately 5.6% of the eligible controls reported that their brothers or father had been affected with prostate cancer.

Genotyping and statistical methods

We genotyped two microsatellite markers consisting of CAG repeats and GGC repeats in exon 1 of the AR gene in all our study subjects. Multiplex polymerase chain reaction (PCR) with fluorescently labeled primers (hex) was performed as described previously (Xu et al. 2001) for both markers by using primers and conditions as described by Irvine et al. (1995). A modified version of the Linkage Designer program (http://dnalab-www.uia.ac.be/dnalab/ld.html) was used to bin the alleles, and inconsistencies were checked by the LINKAGE software (Lathrop et al. 1984; Cottingham et al. 1993) without disease phenotype information. Marker allele frequencies were estimated from the 159 HPC probands.

A linkage disequilibrium (LD) test between CAG and GGC repeats of the AR gene was performed by using the GDA computer program (Weir 1996). The empirical *P*-values of LD tests were based on 10,000 replicate samples. Multipoint linkage analyses were performed by using both parametric and non-parametric methods, implemented by the GENEHUNTER computer program, version 1.3 (Kruglyak et al. 1996). For the parametric analysis, the same genetic model that was used by Xu et al. (1998) was assumed. Linkage in the presence of heterogeneity was assessed by the use of Smith's admixture test for heterogeneity (Ott 1998).

An unconditional logistic regression was used to test for association between genotypes and affection status, adjusting for age. The lengths of CAG and GGC repeats were examined as categorical variables (CAG≤21 vs CAG≥22, and GGC≤16 vs GGC≥17). The categories were defined based on the median value of these two repeats in the controls. Primarily because of the limited sample size of African American and other racial groups, all the analyses were limited to Caucasians only, to decrease the confounding factor of racial differences.

A male-limited X-linked reconstruction-combination transmission/disequilibrium test (XLRC-TDT) was used to test for linkage and association between a prostate cancer susceptibility gene and the length of CAG and GGC repeats (Horvath et al. 2000). XLRC-TDT employs parental-genotyped reconstruction and corrects for the biases resulting from reconstruction. The observed transmission of an allele from a heterozygous mother to affected sons is then compared with the expected rate. We used this method because: (1) prostate cancer is a male-limited disease, (2) AR is on the X-chromosome, and (3) parental genotypes were not available in many families. For our study, the continuity corrected z' statistics were used with corresponding P-values and the exact P-values, which are not dependent on a theoretical large-sample approximation. This method was applied to test for linkage in the presence of association by including the data from multiple sons in a nuclear family. We also used this method to test for association by utilizing the phenotype data of the first son in a nuclear family whose genotype was available. For both of the tests, 159 families were trimmed into 186 independent nuclear families in which mothers are unrelated and informative.

Results

Linkage study of the AR gene region in 159 HPC families

The CAG and GGC repeats of the AR were genotyped in all family members with available DNA samples in the 159 HPC families. Positive linkage scores were observed at the AR region (Table 2), with a maximum LOD under heterogeneity (HLOD) of 0.49 (P=0.12) and a peak NPL Z-score of 0.49 (P=0.30). Stratified linkage analyses based on family characteristics, such as mean age of diagnosis, number of affected members in the family, and ethnicity, were also performed (Table 2). Stronger evidence for linkage at the AR region was observed in the families with a mean age of diagnosis <65 years (n=79, HLOD=1.61, P=0.006) and in the families with ≥5 affected members (n=90, HLOD=0.87, P=0.04). Interestingly, evidence for linkage at the region was primarily from the families ascertained later in our study, as the HLODs in the first 79 families and later 80 families were 0 and 0.68 (P=0.07), respectively. When families were stratified based on the patterns of male-to-male disease transmission, evidence for linkage was provided by families with male-to-male transmission (HLOD=0.72, P=0.07). No evidence for linkage was observed in the subgroup of 60 families with absence of male-to-male transmission (HLOD=0).

Stratified linkage analyses based on probands' genotypes at the CAG or GGC repeats were also performed. Interestingly, among 112 families where probands had ≤16 GGC repeats, we found an HLOD of 0.70 (P=0.07). No evidence for linkage at the AR region, however, was found in 67 families where probands had ≤21 CAG repeats.

Association studies in hereditary and sporadic prostate cancers

In addition to the samples from HPC families, the CAG and GGC repeats of the AR were genotyped in all available sporadic cases and unaffected controls. The two repeats were in strong, but not complete, LD (P=0.0003). The number of CAG repeats ranged from 7 to 34 in the Caucasians. The mean number of CAG repeats was 21.47 (±3.42), 21.80 (±3.32), and 22.02 (±3.15), in controls, sporadic cases, and HPC probands, respectively. The difference in mean CAG repeats among these groups was not

Table 2 Multipoint linkage results in subsets of HPC families (*HLOD* maximum LOD under heterogeneity, *NPL* nonparametric linkage)

Characteristics	No. pedigrees	HLOD		NPL	
		AR-$(CAG)_n$	AR-$(GGC)_n$	AR-$(CAG)_n$	AR-$(GGC)_n$
Overall	159	0.49	0.49	0.49	0.50
Age at diagnosis					
<65	79	1.61	1.61	1.83	1.83
≥65	80	0.00	0.00	–1.14	–1.13
Number of affected members					
=3	29	0.00	0.00	–0.89	–0.88
=4	40	0.00	0.00	–0.19	–0.18
≥5	90	0.87	0.87	1.34	1.34
Ethnicity					
Caucasians	133	0.47	0.47	0.32	0.33
African-Americans	14	0.00	0.00	0.46	0.46
Others	12	0.08	0.08	0.24	0.24
Male-to-male disease transmission					
Male to male	99	0.72	0.73	0.98	0.98
Without male to male	60	0.00	0.00	–0.46	–0.46
AR CAG and GGC repeat length					
CAG ≤21	67	0.00	0.00	–0.85	–0.85
GGC ≤16	112	0.69	0.70	1.08	1.09

Table 3 AR CAG and GGN repeat frequencies in HPC probands, sporadic cases, and unaffected control subjects

Repeat	Frequency (%)					
	HPC probands	Sporadic	Controls	RR[a,b](95% CI)	RR[c] (95% CI)	RR[d] (95% CI)
No. of CAG repeats						
≥22	59 (51)	105 (50)	81 (45)	1	1	1
≤21	57 (49)	105 (50)	99 (55)	0.75 (0.46–1.21)	0.82 (0.55–1.22)	0.81 (0.56–1.17)
No. of GGC repeats						
≥17	37 (29)	63 (32)	72 (41)	1	1	1
≤16	92 (71)	135 (68)	102 (59)	1.69 (1.03–2.78)	1.51 (0.99–2.32)	1.58 (1.08–2.32)
Combined no. of CAG and GGC repeats						
CAG ≥22 and GGC ≥17	10 (9)	32 (17)	30 (18)	1	1	1
CAG ≥22 and GGC ≤16	46 (41)	63 (34)	46 (27)	2.95 (1.28–6.79)	1.27 (0.68–2.39)	1.62 (0.92–2.95)
CAG ≤21 and GGC ≥17	23 (21)	27 (14)	39 (23)	1.76 (0.73–4.26)	0.65 (0.32–1.30)	0.92 (0.49–1.72)
CAG ≤21 and GGC ≤16	33 (29)	66 (35)	54 (32)	1.59 (0.67–3.80)	1.14 (0.62–2.12)	1.29 (0.72–2.29)

[a]All RRs were age adjusted
[b]HPC probands vs controls
[c]Sporadic cases vs controls
[d]All cases vs controls

statistically significant (P=0.56). The number of GGC repeats ranged from 4 to 21 in the Caucasians, with 16 repeats being the most frequent in each of the three groups. The proportion of men with 16 GGC repeats was highest in HPC probands (67%), medium in sporadic cases (60%), and lowest in controls (48%). The difference in the proportion of this allele (16 repeats) among the three groups was marginally significant (P=0.05).

Based on the median repeat length of the controls, the frequency of CAG and GGC repeats were compared as categorical variables (CAG repeats ≤21 or ≥22, and GGC repeats ≤16 or ≥17) between HPC probands, sporadic cases, and controls. Contrary to the results of functional studies, both HPC probands and sporadic cases had lower frequencies of shorter CAG repeats (≤21) than the controls (Table 3), although the results were not significantly different (all Ps>0.2). For the GGC repeats, a higher proportion of men with ≤16 repeats were observed in HPC probands (71%) and sporadic cases (68%) compared with controls (59%). We observed statistically significant differences between HPC probands and controls (P=0.04) and between all prostate cancer cases and controls (P=0.02). The estimated RR for prostate cancer was 1.58 (95% CI=1.08–2.32) for men with ≤16 GGC repeats compared with men who had ≥17 GGC repeats. When the risk of hereditary or sporadic prostate cancer was estimated separately, the risk was higher for hereditary prostate cancer [RR=1.69 (95% CI=1.03–2.78)].

The association between prostate cancer risk and combined genotypes of CAG and GGC repeats was also examined (Table 3). Men with ≥22 CAG and ≥17 GGC repeats (reference group) were compared with men with: (1) ≥22 CAG and ≤16 GGC repeats, (2) ≤21 CAG and ≥17 GGC repeats, or (3) ≤21 CAG and ≤16 GGC repeats (Table 3). Consistent with the results of independent analyses of CAG and GGC repeats in our study, the highest risk for hereditary prostate cancer was observed among men with a genotype of ≥22 CAG repeats and ≤16 GGC repeats (RR=2.95, 95% CI=1.28–6.79).

We examined the relationship of CAG and GGC repeat lengths to Gleason scores and pathological stages among the sporadic prostate cancer cases and also to PSA among the controls. No statistically significant difference was observed in the genotypic frequencies between the groups with low (≤6) versus high (≥7) Gleason scores or between the groups stratified by disease confined to the prostate or non-localized disease. Among the control subjects, we found no statistically significant difference in PSA levels when comparing the men with long or short CAG or GGC repeats (data not shown).

Family-based linkage and association studies

Because TDT in nuclear families is a powerful test for linkage in the presence of association, we performed an XLRC-TDT for AR in 186 independent nuclear families (Table 4). We observed preferential transmission of short GGC alleles (≤16 repeats) from heterozygous mothers to their affected sons (z'=2.65, P=0.008). This increased power for detecting linkage may be explained by the presence of association between the GGC repeats and prostate cancer. A similar result was observed when the 16 GGC repeat was tested (data not shown, z'=3.17, P=0.001). No significant over-transmission of CAG repeat alleles was

Table 4 Results of family-based linkage and association test

Allele	No. informative pedigrees	Z-test		Exact test
		z'	P-values	P-values
Test for linkage in the presence of association				
CAG <=21	74	–1.54	0.12	0.12
GGC <=16	66	2.65	0.008	0.0078
Test for association				
CAG <=21	71	–0.72	0.47	0.47
GGC <=16	65	2.66	0.0071	0.0067

observed. These results strongly suggest GGC repeats are linked to a prostate cancer susceptibility gene.

XLRC-TDT was also used to test for association by using only the first affected son (with available genotype) from each nuclear family (Table 4), thus providing an unbiased association test free of potential population-stratification. Again, mothers who were heterozygous for the ≤16 GGC repeat allele preferentially transmitted the ≤16 GGC repeats to affected sons (z'=2.66, P=0.007). Paralleling the family based XLRC-TDT findings, a similar result was observed when association of the 16 GGC repeat was tested (data not shown, z'=3.05, P=0.002). No significant over-transmission of CAG repeats was observed. These results strongly suggest that the GGC repeats of AR are associated with increased prostate cancer susceptibility.

Discussion

This study evaluated the importance of CAG and GGC repeats in exon 1 of AR by using the following comprehensive approaches: linkage and association analyses, parametric and transmission/disequilibrium tests for linkage, population-based and family-based association tests, and inclusion of both hereditary and sporadic prostate cancer patients. Significant evidence for linkage at the AR region, especially at the GGC locus, was observed in the 159 HPC families (Table 4). Significant association between the length of GGC repeats and prostate cancer risk, especially with hereditary prostate cancer, was observed by using both population-based and family-based association tests (Tables 3, 4). The consistent results between the linkage and association studies increase the confidence in these results and implicate the GGC repeats in AR in prostate cancer susceptibility.

Although prostate cancer linkage at the AR region has been evaluated in another AR-targeted linkage study and in four genome-wide screens, this is the first time that significant evidence has been observed for linkage between prostate cancer susceptibility and a nucleotide repeat marker within the AR gene. The study by Lange et al. (2000) targeted the CAG repeats of AR and found no evidence for linkage. Genome-wide screens are evenly split, with two reporting linkage and two observing no evidence for linkage. However, the absence of linkage in the study by Lange et al. (2000) and in two genome-wide screens does not necessary exclude AR as a potential prostate cancer susceptibility gene. Mixed results from linkage studies can be explained by choice of study design and analytical method, the characteristics of families, and the selection of markers. Our study is different from previous linkage studies in several ways. First, in this study, we employed both traditional parametric linkage analyses and transmission/disequilibrium tests (TDT) for linkage. TDT is a powerful test for linkage of these X-chromosome markers in this male-limited disease: (1) because the majority of mothers are informative for the markers when either directly genotyped or reconstructed from the genotypes of offspring, after correcting for potential bias (Horvath et al. 2000), and (2) because of the presence of association between GGC repeats and prostate cancer. The significant linkage results from XLRC-TDT (P=0.008) and positive, but not statistically significant, results from the traditional parametric linkage analysis (HLOD=0.49, P=0.12) clearly demonstrate the power of the TDT approach. Second, the 159 HPC families included in this study are characterized by higher proportions of early age of diagnosis (50% families with mean age at diagnosis <65 years) and large numbers of affected family members (57% families have five or more affected members). Because these are the characteristics of hereditary families, and because the evidence for linkage has primarily been observed in the families with these characteristics, the overall evidence for linkage in the entire sample depends on the proportion of these families in the study. Last, we used both CAG and GGC repeats within AR in the linkage study. This not only increased the genetic information in the linkage study, but also allowed us to evaluate evidence for linkage in subsets of families stratified by GGC repeats.

The finding that the evidence for linkage is primarily observed in families with male-to-male disease transmission is unexpected. However, this is not completely contradictory to the expectations of X-linkage for a complex disease such as prostate cancer. The grouping of families with or without male-to-male disease transmission is an approximation based on limited information and is subject to misclassification. The potential phenocopies and locus heterogeneity in these families may tend to classify families into the category of male-to-male disease transmission. However, the important point is whether affected men share a maternal allele more often than expected, or whether mothers who are heterozygous for the ≤16 GGC repeats transmitted this allele to affected sons more often than expected, regardless of whether fathers are affected or not. In addition, the stronger evidence for linkage in the families with male-to-male disease transmission may be explained by the hypothesis that AR is a strong modifier gene that works in conjunction with an autosomal susceptibility gene(s). This possibility is consistent with our observation that linkage is primarily provided by the families with a younger mean age of diagnosis (AR accelerates the development of prostate cancer) and by the families with the most affected members (AR assures the penetrance of major susceptibility genes) and is further bolstered by studies that provide linkage evidence for prostate cancer susceptibility genes on several autosomal regions (Smith et al. 1996; Berthon et al. 1998; Gibbs et al. 2000; Berry et al. 2000). Recently, Cui et al. (2001) observed that two locus models, combining autosomal dominant with either an autosomal recessive or X-linked model, fit their data better than did single-locus models in segregation analyses. Their evidence further supports our view that AR is probably a strong modifier gene in the etiology of hereditary prostate cancer.

A significant association between prostate cancer risk and the length of GGC repeats was observed in our population-based association test by comparing the frequency in 159 HPC probands, 245 sporadic cases, and 211 unaf-

fected controls, and also in our family-based association test by comparing the observed and expected transmission of GGC alleles of ≤16 repeats from heterozygous mothers to the first affected son in each nuclear family. The consistent results from population-based and family-based association studies strengthen these results and eliminate the potential bias of population stratification. XLRC-TDT is the correct test for association when one affected son per nuclear family is used and is an informative test in this male-limited X-linked disease because the majority of mothers are "married-in" and thus contribute to the test statistics independently. The stronger evidence for association observed in the family-based association test is also consistent with the higher risk for hereditary prostate cancer observed from our population-based association test. Our study has also demonstrated the importance of studying both hereditary and sporadic patients and of utilizing both population-based and family-based tests.

The significant association of GGC repeats, but not CAG repeats, with prostate cancer risk is consistent with the results from many other association studies. Several studies have failed to observe an association between prostate cancer and CAG repeats (Bratt et al. 1999; Correa-Cerro et al. 1999; Edwards et al. 1999; Lange et al. 2000). For the studies that detected such an association, the statistical evidence is weak and is only observed in certain subgroups (Stanford et al. 1997; Giovannucci et al. 1997). In contrast, although one study failed to detect an association between prostate cancer risk and GGC repeats (Correa-Cerro et al. 1999), several other studies found a significant association (Hakimi et al. 1997; Stanford et al. 1997; Platz et al. 1998). It is possible that weak associations between prostate cancer risk and CAG repeats may reflect the effect of GGC repeats, which are ~1 kb away from, and in LD with, the CAG repeats.

Although we have observed consistent evidence for linkage and association between the GGC repeats of AR and prostate cancer risk, caution should be taken when interpreting and generalizing these findings. The study subjects were recruited primarily for genetics studies rather than for a rigorously designed epidemiological study. Whereas this may enrich the genetic cases in our study sample and help us to identify the risk alleles, the point estimates of RR in this study are difficult to generalize. Furthermore, the control subjects, who were recruited from a prostate cancer screening population, are subject to potential misclassification because they may represent a higher risk population than the general population attributable to self-selection. This potential bias, however, is unlikely to be significant in our study, because very few of the 182 personally interviewed controls reported a positive family history (defined as an affected father and/or brothers). In addition, all control subjects were found to have normal DRE and PSA results at the time of screening.

The combination of linkage and association approaches in our study is an optimal study design in gene mapping of a complex disease such as prostate cancer. In this investigation, we have successfully demonstrated the use of this study design, as this is the first report to provide evidence that linkage at the AR region is from families whose probands have ≤16 GGC repeats. The consistent results from both linkage and association studies strengthen the role of AR as a prostate cancer susceptibility gene. The implication from this study, viz., that the variation of GGC repeats in the AR gene mediates the effect of AR in the etiology of prostate cancer, calls for further studies of this polymorphism in other independent data sets. Functional analysis of the influence of GGC repeat length on AR activity will be a further important step toward understanding the role of this AR polymorphism in prostate cancer risk.

Acknowledgements The authors thank all the study subjects who participated in this investigation and the anonymous reviewers who provided constructive suggestions. This work was partially supported by PHS SPORE CA58236 and two grants from the Department of Defense to W.B.I and J.X.

References

Beilin J, Ball EM, Favaloro JM, Zajac JD (2000) Effect of the androgen receptor CAG repeat polymorphism on transcriptional activity: specificity in prostate and non-prostate cell lines. J Mol Endocrinol 25:85–96

Berry R, Schaid DJ, Smith JR, French AJ, Schroeder JJ, McDonnell SK, Peterson BJ, Wang ZY, Carpten JD, Roberts SG, Tester DJ, Blute ML, Trent JM, Thibodeau SN (2000) Linkage analyses at the chromosome 1 loci 1q24–25 (HPC1), 1q42.2–43 (PCAP), and 1p36 (CAPB) in families with hereditary prostate cancer. Am J Hum Genet 66:539–546

Berthon P, Valeri A, Cohen-Akenine A, Drelon E, Paiss T, Wohr G, Latil A, et al (1998) Predisposing gene for early-onset prostate cancer, localized on chromosome 1q42.2–43. Am J Hum Genet 62:1416–1424

Bratt O, Borg A, Kristoffersson U, Lundgren R, Zhang QX, Olsson H (1999) CAG repeat length in the androgen receptor gene is related to age at diagnosis of prostate cancer and response to endocrine therapy, but not to prostate cancer risk. Br J Cancer. 81:672–676

Carter BS, Beaty TH, Steinberg GD, Childs B, Walsh PC (1992) Mendelian inheritance of familial prostate cancer. Proc Natl Acad Sci USA 89:3367–3371

Chamberlain NL, Driver ED, Miesfeld RL (1994) The length and location of CAG trinucleotide repeats in the androgen receptor N-terminal domain affect transactivation function. Nucleic Acids Res 22:3181–3186

Chang C, Kokontis J, Liao S (1988) Molecular cloning of human and rat complementary DNA encoding androgen receptors. Science 240:324–326

Cerhan JR, et al (1999) Family history and prostate cancer risk in a population-based cohort of Iowa men. Cancer Epidemiol Biomarkers Prev 8:53–60

Coetzee GA, Ross RK (1994) Re: Prostate cancer and the androgen receptor. J Natl Cancer Inst 86:872–873

Correa-Cerro L, Wohr G, Haussler J, Berthon P, Drelon E, Mangin P, Fournier G, Cussenot O, Kraus P, Just W, Paiss T, Cantu JM, Vogel W (1999) $(CAG)_n$ and $(GGC)_n$ repeats in the human androgen receptor gene are not associated with prostate cancer in a French-German population. Eur J Hum Genet 7:357–362

Cottingham RW Jr, Idury RM, Schaffer AA (1993) Faster sequential genetic linkage computations. Am J Hum Genet 53:252–263

Crocitto LE, Henderson BE, Coetzee GA. (1997) Identification of two germline point mutations in the 5'UTR of the androgen receptor gene in men with prostate cancer. J Urol 158:1599–1601

Cui J, Staples MP, Hopper JL, English DR, McCredie MR, Giles GG (2001) Segregation analyses of 1,476 populataion-based Australian families affected by prostate cancer. Am J Hum Genet 68:1207–1218

Dunnen JT den, Antonarakis SE (2001) Nomenclature for the description of human sequence variations. Hum Genet 109:121–124

Edwards A, Hammond HA, Jin L, Caskey CT, Chakraborty R (1992) Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. Genomics 12: 241–253

Edwards SM, Badzioch MD, Minter R, Hamoudi R, Collins N, Ardern-Jones A, Dowe A, Osborne S, Kelly J, Shearer R, Easton DF, Saunders GF, Dearnaley DP, Eeles RA (1999) Androgen receptor polymorphisms: association with prostate cancer risk, relapse and overall survival. Int J Cancer 84:458–465

Elo JP, Kvist L, Leinonen K, Isomaa V, Henttu P, Lukkarinen O, Vihko P (1995) Mutated human androgen receptor gene detected in a prostatic cancer patient is also activated by estradiol. J Clin Endocrinol Metab 80:3494–3500

Gibbs M, Stanford JL, Jarvik GP, Janer M, Badzioch M, Peters MA, Goode EL, Kolb S, Chakrabarti L, Shook M, Basom R, Ostrander EA, Hood L (2000) A genomic scan of families with prostate cancer identifies multiple regions of interest. Am J Hum Genet 67:100–109

Giovannucci E, Stampfer MJ, Krithivas K, Brown M, Dahl D, Brufsky A, Talcott J, Hennekens CH, Kantoff PW (1997) The CAG repeat within the androgen receptor gene and its relationship to prostate cancer. Proc Natl Acad Sci USA 94:3320–323

Goddard KA, Witte JS, Suarez BK, Catalona WJ, Olson JM (2001) Model-free linkage analysis with covariates confirms linkage of prostate cancer to chromosomes 1 and 4. Am J Hum Genet 68:1197–206

Gronberg H, Damber L, Damber JE, Iselius L. (1997) Segregation analysis of prostate cancer in Sweden: support for dominant inheritance. Am J Epidemiol 146:552–557

Hakimi JM, Schoenberg MP, Rondinelli RH, Piantadosi S, Barrack ER (1997) Androgen receptor variants with short glutamine or glycine repeats may identify unique subpopulations of men with prostate cancer. Clin Cancer Res 3:1599–1608

Hayes RB, Liff JM, Pottern LM, Greenberg RS, Schoenberg JB, Schwartz AG, Swanson GM, Silverman DT, Brown LM, Hoover RN, et al (1995) Prostate cancer risk in US blacks and whites with a family history of cancer. Int J Cancer 60:361–364

Horvath S, Laird NM, Knapp M (2000) The transmission/disequilibrium test and parental-genotype reconstruction for X-chromosomal markers. Am J Hum Genet 66:1161–1167

Hsieh CI, Oakley-Girvan I, Balise RR, Halpern J, Gallagher RP, Wu AH, Kolonel LN, O'Brien LE, Lin IG, Van Den Berg DJ, Teh CZ, West DW, Whittemore AS (2001) A genome screen of families with multiple cases of prostate cancer: evidence of genetic heterogeneity. Am J Hum Genet 69:148–158

Hsing AW, Gao YT, Wu G, Wang X, Deng J, Chen YL, Sesterhenn IA, Mostofi FK, Benichou J, Chang C (2000) Polymorphic CAG and GGN repeat lengths in the androgen receptor gene and prostate cancer risk: a population-based case-control study in China. Cancer Res 60:5111–5116

Ingles SA, Ross RK, Yu, MC, Irvine RA, La Pera G, Haile RW, Coetzee GA (1997) Association of prostate cancer risk with genetic polymorphisms in vitamin D receptor and androgen receptor. J.Natl Cancer Inst 89:166–170

Irvine RA, Yu MC, Ross RK, Coetzee GA (1995) The CAG and GGC microsatellites of the androgen receptor gene are in linkage disequilibrium in men with prostate cancer. Cancer Res 55: 1937–1940

Knapp M (1999) The transmission/disequilibrium test and parental-genotype reconstruction: the reconstruction-combined transmission/disequilibrium test. Am J Hum Genet 64:861–870

Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES (1996) Parametric and nonparametric linkage analysis: a unified multipoint approach. Am J Hum Genet 58:1347–1363

Lange EM, Chen H, Brierley K, Livermore H, Wojno KJ, Langefeld CD, Lange K, Cooney KA (2000) The polymorphic exon 1 androgen receptor CAG repeat in men with a potential inherited predisposition to prostate cancer. Cancer Epidemiol Biomarkers Prev 9:439–442

Lathrop GM, Lalouel JM, Julier C, Ott J (1984) Strategies for multilocus linkage analysis in humans. Proc Natl Acad Sci USA 81:3443–3446

Lubahn DB, Joseph DR, Sar M, Tan J, Higgs HN, Larson RE, French FS, Wilson EM (1988) The human androgen receptor: complementary deoxyribonucleic acid cloning, sequence analysis and gene expression in prostate. Mol Endocrinol 2:1265–1275

Mhatre AN, Trifiro MA, Kaufman M, Kazemi-Esfarjani P, Figlewicz D, Rouleau G, Pinsky L (1993) Reduced transcriptional regulatory competence of the androgen receptor in X-linked spinal and bulbar muscular atrophy. Nat Genet 5:184–188

Miller EA, Stanford JL, Hsu L, Noonan E, Ostrander EA (2001) Polymorphic repeats in the androgen receptor gene in high-risk sibships. Prostate 48:200–205

Mononen N, Syrjakoski K, Matikainen M, Tammela TL, Schleutker J, Kallioniemi OP, Trapman J, Koivisto PA (2000) Two percent of Finnish prostate cancer patients have a germ-line mutation in the hormone-binding domain of the androgen receptor gene. Cancer Res 60:6479–6481

Monroe KR, Yu MC, Kolonel LN, Coetzee GA, Wilkens LR, Ross RK, Henderson BE (1995) Evidence of an X-linked or recessive genetic component to prostate cancer risk. Nat Med 1:827–829

Narod SA, Dupont A, Cusan L, Diamond P, Gomez JL, Suburu R, Labrie F (1995) The impact of family history on early detection of prostate cancer. Nat Med 1:99–101

Ott J (1998) Analysis of human genetic linkage, 3rd edn. Johns Hopkins University Press, Baltimore

Platz EA, Giovannucci E, Dahl DM, Krithivas K, Hennekens CH, Brown M, Stampfer MJ, Kantoff PW (1998) The androgen receptor gene GGN microsatellite and prostate cancer risk. Cancer Epidemiol Biomarkers Prev 7:379–384

Schaid DJ, McDonnell SK, Blute ML, Thibodeau SN (1998) Evidence for autosomal dominant inheritance of prostate cancer. Am J Hum Genet 62:1425–1438

Schuurman AG, Zeegers MP, Goldbohm RA, Brandt PA van den (1999) A case-cohort study on prostate cancer risk in relation to family history of prostate cancer. Epidemiology 10:192–195

Sleddens HF, Oostra BA, Brinkmann AO, Trapman J (1993) Trinucleotide (GGC) repeat polymorphism in the human androgen receptor (AR) gene. Hum Mol Genet 2:493

Smith JR, Freije D, Carpten JD, Gronberg H, Xu J, Isaacs SD, Brownstein MJ, Bova GS, Guo H, Bujnovszky P, Nusskern DR, Damber JE, Bergh A, Emanuelsson M, Kallioniemi OP, Walker-Daniels J, Bailey-Wilson JE, Beaty TH, Meyers DA, Walsh PC, Collins FS, Trent JM, Isaacs WB (1996) Major susceptibility locus for prostate cancer on chromosome 1 suggested by a genome-wide search. Science 274:1371–1374

Spielman RS, McGinnis RE, Ewens WJ (1993) Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). Am J Hum Genet 52: 506–516

Stanford JL, Just JJ, Gibbs M, Wicklund KG, Neal CL, Blumenstein BA, Ostrander EA (1997) Polymorphic repeats in the androgen receptor gene: molecular markers of prostate cancer risk. Cancer Res 57:1194–1198

Weir BS (1996) Genetic data analysis II: methods for discrete population genetic data. Sinauer, Sunderland

Woolf CM (1960) An investigation of familial aspects of carcinoma of the prostate. Cancer 13:739–744

Xu J, Meyers D, Freije D, Isaacs S, Wiley K, Nusskern D, Ewing C, Wilkens E, Bujnovszky P, Bova GS, Walsh P, Isaacs W, Schleutker J, Matikainen M, Tammela T, Visakorpi T, Kallioniemi OP, Berry R, Schaid D, French A, McDonnell S, Schroeder J, Blute M, Thibodeau S, Trent J, et al (1998) Evidence for a prostate cancer susceptibility locus on the X chromosome. Nat Genet 20:175–179

Xu J, Zheng SL, Chang B, Smith JR, Carpten JD, Stine OC, Isaacs SD, Wiley KE, Henning L, Ewing C, Bujnovszky P, Bleeker ER, Walsh PC, Trent JM, Meyers DA, Isaacs WB (2001) Linkage of prostate cancer susceptibility loci to chromosome 1. Hum Genet 108:335–345

ELSEVIER

Cancer Genetics and Cytogenetics 137 (2002) 1–7

CANCER GENETICS AND CYTOGENETICS

Lead article

Germline sequence variants of the *LZTS1* gene are associated with prostate cancer risk

Gregory A. Hawkins[a], Josyf C. Mychaleckyj[a], Siqun L. Zheng[a], Dennis A. Faith[b], Brian Kelly[b], S.D. Isaacs[b], Kathleen E. Wiley[b], Bao-Li Chang[a], Charles M. Ewing[b], Piroska Bujnovszky[b], Eugene R. Bleecker[a], P.C. Walsh[b], Deborah A. Meyers[a,*], William B. Isaacs[b], Jianfeng Xu[a]

[a]*Center for Human Genomics, Wake Forest University School of Medicine, Winston-Salem, NC, USA*
[b]*Department of Urology, Johns Hopkins Medical Institutions, Baltimore, MD, USA*

Received 6 November 2001; received in revised form 14 February 2002; accepted 15 February 2002

Abstract The 8p22∼p23 region has been identified as a potential site for genes associated with prostate cancer. The gene *LZTS1* has been mapped to the 8p22∼p23 region and identified as a potential tumor suppressor based on loss of heterozygosity studies using primary esophageal tumors. Sequence analysis of mRNA from various tumors has revealed multiple mutations and aberrant mRNA transcripts. The most recent report associates *LZTS1* function with stabilization of $p34^{cdc2}$ during the late S-G_2/M stage of mitosis, affecting normal cell growth. In this study, a detailed DNA sequence analysis of *LZTS1* was performed in a screening panel consisting of sporadic and hereditary prostate cancer (HPC) cases and unaffected controls. Twenty-four SNP, 15 of which were novel, were identified in germline DNA. Four coding SNP were identified. Eleven informative SNP were genotyped in 159 HPC probands, 245 sporadic prostate cancer cases, and 222 unaffected controls. Four of these SNP were statistically significant for association with prostate cancer ($P \leq 0.04$). These results add evidence supporting a role of *LZTS1* in prostate cancer risk.

1. Introduction

Loss of heterozygosity (LOH) in the 8p arm is a common characteristic of numerous types of cancer including prostate cancer [1]. In the case of prostate cancer, allelic loss >60% has been observed at 8p22 in some prostate tumors [2,3]. Consistent with the LOH, evidence for linkage between a prostate cancer susceptibility gene and markers at 8p22∼p23 has been observed in several studies [4–6]. Based on these observations, it is possible that the same tumor suppressor genes (TSG) in this region may be the target of function-altering genomic changes both somatically and in the germline, leading to an increased risk for prostate cancer.

A number of genes have now been physically mapped to the 8p22∼p23 region and investigated for TSG activity [7–9]. One such gene, *LZTS1* (leucine zipper tumor suppressor, originally termed FEZ1), has been identified as a potential TSG [10]. *LZTS1* encodes a 596 amino acid protein (67 kDa) from a 6.8 kb transcript and was identified during an LOH study using primary esophageal cancer samples. *LZTS1* has been physically mapped (YAC and BAC contigs) to an ∼2.5 Mb region on 8p22 between the STS markers D8S1715 and D8S258 and lies proximal to the LPL loci [11].

The initial analysis of *LZTS1* protein sequence revealed a short leucine-zipper motif and a 32% sequence identity to a cAMP-responsive activating-transcription factor (Atf5), which suggested the *LZTS1* gene as a potential DNA transcription regulator [11]. The *LZTS1* gene is ubiquitously expressed in all normal tissues tested (including prostate), but has its highest expression level in testes. Five aberrant mRNA transcripts for *LZTS1* have been detected. Extensive mRNA analysis of *LZTS1* in 41 tumor cell lines (breast, prostate, esophageal, cervical, leukemia, Burkitt lymphoma, colorectal, lung, and melanoma) and 25 primary tumors (prostate, esophageal, and breast) found the absence of expression in 76% of tumor cell lines and 64% of primary tumor cells. Truncated *LZTS1* mRNA was also detected in prostate, colorectal, esophageal, leukemia, and melanoma tumor cells. In an additional study by Vecchione et al. [12], aberrant expression of *LZTS1* in gastric carcinoma was detected and correlated to LOH of flanking markers D8S261 and LPL.

* Corresponding author. Tel.: 336-716-5700; fax: 336-716-5777.

0165-4608/02/$ – see front matter
PII: S0165-4608(02)00549-6

Sequence analysis of exons 1–3 of 194 cancers (tumors and cell lines from esophageal, prostate, breast, ovarian, leukemia, cervical [10], and 26 gastric carcinomas [12]) revealed point mutations that caused either structural changes or truncation of the protein product in two primary esophageal tumors (S29P and K119E), in the PC3 prostate cancer cell line (Q501Ter) [10], and one gastric carcinoma (H17A) [12]. This observation for cell line PC3 is notable since the Gln501Ter mutation would complement the 8p arm deletion that characterizes PC3, making *LZTS1* a potential classic two-hit tumor suppressor gene. Southern blot analysis of genomic DNA from esophageal, prostate, and breast cancer cell lines found only one case of possible genomic rearrangement in the breast cancer cell line MB436S [10]. In the report by Cabeza-Arvelaiz et al. [11], transfection of the complete *LZTS1* gene into rat and human prostate cancer cell lines was shown to suppress tumor cell colony growth, adding evidence that *LZTS1* is silenced in these prostate cancer cell lines and is indeed a tumor suppressor gene. The most recent report confirms the tumor suppression activity of *LZTS1* and presents convincing data that this activity is probably influenced by *LZTS1* stabilization of $p34^{cdc2}$ during the late S-G_2/M stage of mitosis, which subsequently affects normal cell growth [13].

All of the findings presented above strongly support the role of *LZTS1* in tumor development, however, these studies have been performed studying somatic alterations only. Based on evidence listed above and the fact that *LZTS1* gene lies near the 8p22 region, we performed a detailed DNA sequence analysis of *LZTS1* in a screening panel of 96 subjects consisting of sporadic and hereditary prostate cancer (HPC), as well as unaffected controls to determine if previously published SNP are present in germline DNA and to identify any new genetic variants within *LZTS1*. We then tested several informative SNP within a target region of the *LZTS1* gene in 159 HPC probands, 245 sporadic prostate cancer cases, and 222 unaffected controls to determine if any association exists between the SNP and prostate cancer risk.

2. Materials and methods

2.1. Study subjects

A detailed description of the study samples was presented elsewhere [4]. Briefly, a total of 159 HPC probands were ascertained at the Brady Urology Institute of Johns Hopkins Hospital (Baltimore, MD, USA), through referrals, medical records of patients seen at Johns Hopkins Hospital for treatment of prostate cancer, and respondents to various lay publications describing our studies. All of the probands had at least two first-degree relatives affected with prostate cancer. Medical records verified a diagnosis of prostate cancer. The mean age at prostate cancer diagnosis for these probands was 61 years. Among the probands, 133 (84%) were Caucasians and 14 (8.8%) were African-Americans.

Two hundred and forty-five unrelated prostate cancer cases were recruited from patients who underwent treatment for prostate cancer at the John Hopkins Hospital and did not have first-degree relatives affected with prostate cancer. The diagnosis of prostate cancer for all these subjects was confirmed by pathology reports. Preoperative prostate specific antigen (PSA) levels, Gleason score, and pathologic stages were available for 202, 240, and 241 cases, respectively. Mean age at diagnosis for these cases was 58.7 years. Over 93% of the cases were Caucasians, and 3.2% were African-Americans.

Two hundred and twenty-two nonprostate cancer controls were selected from men participating in screening programs for prostate cancer. By applying the exclusion criteria of abnormal digital rectal examination (DRE) and abnormal PSA level (i.e., $\geq$4 ng/ml), 211 were eligible for the study. The mean age at examination was 58 years. Over 86% of the eligible controls were Caucasians and 7.1% were African-Americans. About 5.6% of the eligible controls have brothers or father affected with prostate cancer. The affection status of relatives was obtained by interview of the probands.

2.2. Statistical methods

Hardy-Weinberg Equilibrium (HWE) tests for all SNP, and linkage disequilibrium (LD) tests for all pairs of SNP, were performed using the GDA computer program [14]. Exact HWE tests were performed by permuting the alleles among genotypes and computing the proportion with a smaller conditional probability than the original data. The LD tests were based on an exact test assuming multinomial probability of the multi-locus genotype, conditional on the single-locus genotype [15]. A Monte Carlo simulation was used to assess the significance, by permuting the single-locus genotypes among individuals in the sample to simulate the null distribution. The empirical p-values of both the HWE and LD tests were based on 10,000 replicate samples.

Tests for associations between the SNP and prostate cancer were performed by comparing allele and genotype frequencies between cases and controls. Allele frequencies were estimated by direct count. The hypotheses of differences in allele frequencies between cases and controls were tested based on the χ^2 of Amitage trend tests [16], adjusting for age.

Haplotype frequencies in unrelated individuals were estimated by maximum likelihood estimation, using the best state of haplotype composition (http://www.bioinf.mdc-berlin.de/hap/ithap-help.html). We assumed equal prior probabilities as a starting point for the expectation maximization (EM) algorithm.

2.3. Polymerase chain reaction (PCR) amplification of target regions

For SNP discovery, a screening panel consisting of 96 Caucasian and African-American DNA samples was con-

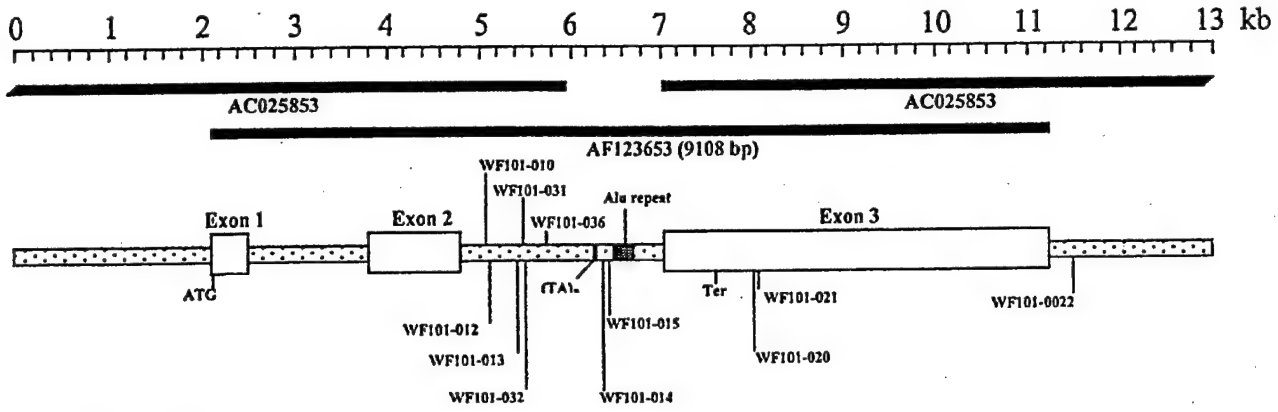


Fig. 1. Graphical representation of the genomic structure of *LZTS1*. The precise exon-boundaries were determined informatically by optimal alignment of accession AF123653 with reference human genomic sequence accession AC025853. The coding regions of the gene are shaded in cross hatch, and the start ATG and stop codons are indicated.

structed from the control, sporadic, and hereditary prostate cancer (HPC) DNA sets. The panel consisted of 32 sporadic cases (24 Caucasian and 8 African-American), 32 HPC samples (24 Caucasian and 8 African-American) and 32 control samples (24 Caucasian and 8 African-American). Our strategy for SNP verification and discovery was to generate 400–600 bp overlapping PCR products covering approximately 7 kb of *LZTS1*, including the complete coding region, sections of intron 1 that flank exons 1 and 2, all of intron 2, and selected portions of the 5′UTR and 3′UTR regions (Fig 1).

The PCR and sequencing primers were derived from the *LZTS1* genomic sequence (GenBank accession AF123653). Additional sequence alignments and flanking sequence were derived from a large genomic clone (GenBank accession # AC025853) and a cDNA clone (GenBank accession #AF123659). We had difficulty amplifying the region surrounding the $(TA)_n$ repeat in intron 2 after multiple attempts and with different primer pairs. To overcome this problem, an ~3500 bp fragment containing all of intron 2 was generated and sequence flanking the $(TA)_n$ repeat was determined using internal sequencing primers. When possible, multiple SNP assays were performed from a single PCR product to reduce the number of sequencing reactions required for genotyping and to maintain the integrity of SNP haplotype determination. Table 1 lists the primer sequences and relative region of *LZTS1* amplified by each primer pair.

Each 30 μL PCR contained 30 ng of genomic DNA, 1 X PCR buffer (Life Technologies, Gaithersburg, MD, USA), 1.5 mM $MgCl_2$, 200 μM dNTP, 15 pmoles of each forward and reverse primer, and 0.5 U of Taq polymerase (Life Technologies). Depending on prior optimization conditions, general cycling conditions were; 94°C for 4 minutes, followed by 25–30 cycles of 94°C for 1 minute, T_{anneal} for 1 minute, and 72°C for 1 minute, and finishing with a single extension cycle of 72°C for 5 minutes. A random sampling of eight reactions was analyzed on a 2% agarose gel. The PCR products were purified using the Quickstep 96 well PCR purification kit (Edge Biosystems, Gaithersburg, MD, USA) and stored in water at −20°C.

2.4. DNA sequencing and allele genotyping

DNA sequencing was performed using the ABI BigDye Terminator sequencing kit (Applied Biosystems, Foster City, CA, USA). Each 10 μL sequencing reaction contained 10–50 ng of purified PCR product, 1.5 pmoles of sequencing primer, 1 μL of BigDye Terminator mix, 1.5 μL of 5× sequencing dilution buffer (400 mM Tris pH 9.0 and 10 mM $MgCl_2$), and water to volume. Cycling conditions were 94°C for 1 minute, 25 cycles of 94°C for 30 seconds, 50°C for 30 seconds, 60°C for 4 minutes, and finishing with a single 72°C extension step for 5 minutes. Sequencing products were ethanol precipitated, air-dried, resuspended in 25 μL ddH_20, and analyzed on an ABI 3700 DNA Analyzer. DNA sequencing data was aligned and polymorphisms identified using Sequencher DNA analysis software (Gene Codes Corporation, Ann Arbor, MI, USA).

3. Results

3.1. Analysis of LZTS1 DNA Sequence for SNP

A total of 24 SNP were identified in this study. Fifteen of the 24 SNP were novel (Table 2). Each SNP was assigned a Wake Forest University Genome Center identifier number (WF101-XXX) for the purpose of this study. Four SNP were found in exons, three of which are novel. One SNP (WF101–016) occurred in exon 3, resulting in the nonsynonymous change Ala461Val, but only in one control sample. The novel SNP WF101–008 and WF101–009 occurred in exon 2, resulting in synonymous changes (Asp259Asp and Glu267Glu, respectively) in both Caucasian and African-American samples. One SNP ascertained from the NCBI dbSNP database, WF101–017 [rs723874], occurred in exon 3 at low frequencies in both Caucasians and African-Americans and caused the amino acid change Leu475Val. WF101–009 was the only SNP in an exon present at a frequency >5%.

The remaining 20 SNP were located in noncoding sequence. Twelve of these SNP were novel. A search of

Table 1
PCR and sequencing primers

Primer ID	Forward	Reverse	Region sequenced	Annealing temperature (°C)
1	AAAAATGGGGTTTCTCTAAGTTGC	AACAAACATTCACTGGGAGCC	Promoter	60
2	TTGGTTTGCTTCTGGCCTCT	GACTCGGGGCTGAGGATG	Promoter/exon 1	60
3	CTTGCTGCCACAGCCTTTC	TGGACGGGTCAAAGTCCAC	Exon 1	60
4	TCACCCAGATTACACGGCAC	CAGATGAGAACAGGGCTCCC	Exon 1/intron 1	60
5	GTGCCCAAGTCGCCATTAG	AGTCTGACAGCGCCCCAG	Intron 1/exon 2	60
6	GCCATCCTGCACTCCTCC	ACCGCTGACCACCCAAAC	Exon 2/intron 2	Ramped 72-62
7	ATGAAGCCGGAAGCCAGAT	CAGGCTGACACCAAAACCAA	Intron 2	60
8	AGGCTGGAATGCCTAACACC	TTCTGGTACTGAATCACCTTCTCC	Intron 2/exon 3	Ramped 68-58
9	CTGGGAAAGCCAGAGGAGT	GATTCAGTACCAGAAACAGCTGC	Exon 3/3′ UTR	60
10	GGAGCCCTTGGAGGTTGAC	TCTGATGGGGCTGGTTCC	3′ UTR	60
11	GGACCTTATCTGTGAAATGAGAGG	TTTGTCCCAAAGCTGGGG	Non coding region	60
12	AGCAAATGTATGGCTTGGCA	GTGCCTTGAGACCCAGGATC	Intron 2/exon 3	62
13	TTCAGCAGGAGAAGCGGC	CTAGTGGGTGCAGTCCCTCC	Intron 2	64
14	ACCCCCAGTTCAGAGTCCAA	ATCCTGGGTAGGGTCGGATT	Intron 2	62
15	TAAGAGTGAATGAAGTCCAGAGCA	TGCCAAGCCATACATTTGCT	Intron 2	60
16	CCTGGCTTTGAACCTCAGCT	CCCAGGTTATCGAGCTAGGC	Intron 2	60
17	GCGCAAGAAGAACGAGGC	AAGCCAGAGGAGTCAGGGC	Exon 3/3′UTR	62
18	GCCATCCTGCACTCCTCC	TTCTGGTACTGAATCACCTTCTCC	Exon 2/intron 2/exon 3	60

dbSNP database had revealed ten predicted SNPs in the noncoding regions of *LZTS1*. We found eight of the 10 predicted SNP in our screening panel, with only WF101–018 [rs904004] and WF101–019 [rs904003] undetected. Fourteen of the nonnoncoding SNP occurred at frequencies >5% in Caucasians. Ishii et al. [10] and Vecchione et al. [12] had previously reported four SNP in *LZTS1* exons (WF101–041, –042, –043, and –044) based on screening DNA isolated from various tumor cells and cell lines. We did not detect these SNP in either Caucasians or African-Americans in our screening panel, suggesting that these SNP are population or carcinoma specific. Our sequence analysis also confirmed the presence of an Alu-Sx repetitive element (positions 4271–4541, E = 2e − 70, 91%) 5′ of exon 3.

In the report by Ishii et al. [10], five aberrant *LZTS1* transcripts were described. In all five cases the aberrant transcripts contained deletions in or near the 5′ end of exon 3. In addition to the exon 3 deletions, one transcript also had a large 3′ portion of exon 2 deleted, whereas another transcript had a 3′ portion of exon 1 and all of exon 2 deleted. In all five cases, the 3′ end of exon 3 was intact. Based this information and our sequencing showing that SNP in the coding region are not predicted to greatly affect the *LZTS1* product, it seemed more likely that defects in DNA transcription of *LZTS1* may have an important role in prostate cancer. As a result of these findings, we chose to focus on 11 informative SNP that lay in or near the exon 2-intron 2-exon 3 region for our association study.

3.2. Analysis of SNP association

Eleven informative SNP in the exon 2-intron 2-exon 3 region (Fig. 1) were genotyped in sets of 159 HPC probands, 245 sporadic prostate cancer cases, and 222 unaffected controls. To decrease the potential impact of population stratification, all of the following analyses were limited to Caucasians. All of the SNP were in HWE in the HPC probands, sporadic cases, and controls. Pair-wise LD tests for all SNP were also performed and they were all in strong LD (all P values $< 10^{-5}$).

Nominally significant higher frequencies of the A allele of WF101–010, C allele of WF101–012, C allele of WF101–031, and G allele of WF101–014 were observed in sporadic cases compared with controls (Table 3). These four specific SNP alleles were also more frequent in HPC probands compared with controls, although not statistically significant. Analysis of the haplotype frequencies based on these four SNP indicated higher haplotype frequencies of A/C/C/G alleles of SNP WF101–010, WF101–012, WF101–031, and WF101–014 in sporadic cases (45%) and in HPC probands (40%), compared with the unaffected controls (36%). The difference is not statistically significant.

4. Discussion

In all previous studies of *LZTS1*, speculation over whether *LZTS1* has any role in tumor suppression was based on functional and genetic analysis from tumors and tumor cell lines. In this report, we present the first association study of *LZTS1* in a case/control/HPC population, based on detailed SNP analysis of the *LZTS1* gene. After resequencing the *LZTS1* gene in a case/control/HPC screening panel, 24 SNP were detected. Of the SNP previously found in tumors and tumor cells [10,12], none were found in the *LZTS1* genomic sequence during this study, suggesting that these genetic variants did not originate in germline cells but were caused by mutational events during tumor formation. Four SNP, three of which are new, were found in the coding region. However, each of these SNP either created a silent mutational change or had a frequency in the coding region

Table 2
LZTS1 SNP and allele frequencies

Identifier	Source	Variant[b]	Amino acid change	Frequency of rare allele[a]					
				Caucasian			African-American		
				Control	Case	HPC	Control	Case	HPC
WF101-001	New	−337G → A		0 (44)	0 (46)	0 (48)	4 (16)	2 (16)	0 (16)
WF101-002	New	−259C → G		0 (46)	3 (46)	1 (48)	0 (14)	0 (16)	0 (16)
WF101-003	New	−132C → G		0 (46)	1 (46)	3 (48)	1 (14)	0 (16)	4 (16)
WF101-004	New	−129C → T		0 (46)	0 (46)	1 (48)	0 (14)	0 (16)	0 (16)
WF101-041	[12]	50A → G	CAC → CGC: H17R	0 (46)	0 (46)	0 (48)	0 (16)	0 (16)	0 (16)
WF101-042	[10]	85T → C	TCC → CCC: S29P	0 (46)	0 (46)	0 (48)	0 (16)	0 (16)	0 (16)
WF101-005	New	1406T → C		12 (44)	12 (46)	14 (48)	7 (14)	4 (14)	6 (16)
WF101-006	New	1524C → T		0 (46)	2 (46)	1 (46)	0 (14)	1 (14)	0 (16)
WF101-007	New	1552C → T		2 (46)	3 (46)	1 (46)	0 (14)	0 (14)	0 (16)
WF101-043	[10]	1605A → G	AAG → GAG: K119E	0 (46)	0 (46)	0 (48)	0 (16)	0 (16)	0 (16)
WF101-008	New	2027C → T	GAC → GAT: D259D	1 (46)	1 (36)	3 (46)	1 (16)	0 (12)	0 (14)
WF101-009	New	2051G → A	GAA → GAG: E267E	14 (46)	9 (36)	14 (26)	8 (16)	3 (12)	5 (14)
WF101-010	rs904000	2812G → A			[c]		2 (16)	1 (12)	0 (12)
WF101-012	rs903999	2883T → C			[c]		3 (16)	2 (12)	1 (12)
WF101-013	rs903998	3200A → C			[c]		2 (16)	5 (12)	4 (10)
WF101-031	New	3329C → T			[c]		8 (16)	8 (16)	6 (10)
WF101-032	New	3338C → C			[c]		2 (16)	5 (16)	2 (10)
WF101-035	New	3505G → T			[c]		8 (16)	8 (14)	6 (10)
WF101-014	New	4361C → T			[c]		3 (16)	2 (14)	1 (14)
WF101-015	New	4544T → A			[c]		8 (16)	4 (14)	4 (14)
WF101-016	New	5033C → T	GCG → GTG: A461V	1 (40)	0 (44)	0 (46)	0 (16)	0 (16)	0 (12)
WF101-017	rs723874	5084C → G	CTG → GTG: L475V	2 (40)	1 (44)	0 (46)	1 (16)	0 (16)	2 (14)
WF101-044	[10]	5152C → T	CAG → TAG: Q501Ter	0 (46)	0 (46)	0 (48)	0 (16)	0 (16)	0 (16)
WF101-018	rs904004	5439C → T	ATC → ATT: I596I	0 (46)	0 (46)	0 (48)	0 (16)	0 (16)	0 (16)
WF101-019	rs904003	5532T → C		0 (46)	0 (46)	0 (48)	0 (16)	0 (16)	0 (16)
WF101-020	rs732337	5785T → A			[c]		2 (16)	1 (10)	N/D
WF101-021	rs904002	5801C → T			[c]		0 (16)	0 (10)	N/D
WF101-022	rs904001	9405C → T			[c]		0 (16)	3 (10)	N/D
WF101-029	rs221894	9907C → T		6 (32)	6 (42)	0 (8)	0 (12)	1 (8)	N/D
WF101-028	rs221893	9938C → A		5 (32)	6 (42)	0 (8)	1 (12)	1 (8)	N/D

[a]Number of chromosomes sequenced in parentheses.

[b]All positions are quoted for reference genomic sequence GenBank accession no. AF123653. The first base of the initial methionine ATG codon is designated +1, and base immediately preceding the start codon designated as −1, as per convention.

[c]See Table 3.

Abbreviations: HPC, hereditary prostate cancer; N/D, not determined.

that was too low to be important in our population. Of the remaining SNP, the 11 that were more common were typed in a larger case/control/HPC set.

The statistically significant associations observed between prostate cancer risk and four SNPs could potentially be due to any of a combination of the following factors. First, the sequence variants themselves could increase the risk of prostate cancer through some unknown mechanisms. Second, the SNP may not increase the risk of prostate cancer, but could be in LD with unknown sequence variants in this region that increase the risk to prostate cancer. Third, the differences in the allele frequencies between cases and controls could be due to different genetic backgrounds in cases and controls. (i.e., population stratification). We attempted to lessen the impact of this source of population stratification by limiting our analyses to Caucasian men only, although this approach might not fully remove the potential impact. On the other hand, based on a sample of 24 consecutive SNP on chromosomes 1, 8, 11, 12, and X that were recently genotyped in this population, we found no evidence to suggest that population stratification exists within our Caucasian case and control samples (data not shown). A family-based association test could be an alternative study design that may overcome the potential bias of population stratification. However, a family based association study would be inefficient in this population because most parents of affected men would most likely be deceased due to the late age of onset of prostate cancer. Finally, the significant associations may be due to the multiple tests performed in this study. Not only were multiple SNP genotyped, but multiple groups (HPC proband and sporadic cases) were also tested for each SNP. Using the commonly suggested Bonferoni test, we calculated adjusted significance levels by multiplying the nominal P values by the total number of tests performed in the study ($N = 24$). After the adjustment, no significant difference was found.

Two other caveats of our study are worth noting. First, the study subjects were recruited primarily for genetics

Table 3
Allele frequencies of the sequence variants in *LZTS1*

SNP	Risk allele	Frequencies			*P* values (compared to controls)[a]	
		Controls	Sporadic	HPC	Sporadic	HPC
WF101-010	A	0.37	0.46	0.41	0.03	NS
WF101-012	C	0.39	0.48	0.46	0.04	NS
WF101-013	A	0.79	0.81	0.81	NS	NS
WF101-031	C	0.59	0.66	0.66	0.03	NS
WF101-032	C	0.20	0.20	0.21	NS	NS
WF101-036	G	0.64	0.68	0.69	NS	NS
WF101-014	G	0.37	0.45	0.43	0.04	NS
WF101-015	T	0.66	0.71	0.69	NS	NS
WF101-020	T	0.37	0.45	0.47	NS	NS
WF101-021	C	0.78	0.80	0.83	NS	NS
WF101-022	C	0.78	0.82	0.79	NS	NS

[a] *P* values were based on the Chi-square of the Amitage trend tests and adjusted for age. Caucasians only.

Abbreviation: NS, not significant.

studies rather than a rigorously designed epidemiologic study. Thus it is difficult to generalize these findings to the general population. However, this study does provide some valuable results. The SNP identified in our studies may be useful in future studies of prostate cancer and other diseases. The increased frequencies of variant alleles at several SNP in the cases should prompt further studies. Second, the source of our control subjects, which were recruited from a prostate cancer screening population, is questionable. This control group may represent a higher risk population compared to the general population due to self-selection. This potential bias, however, is unlikely to be significant in our study. All control subjects were found to have a normal digital rectal examination (DRE) and PSA results at the time of screening. Furthermore, very few of the 182 controls interviewed reported a positive family history (defined as an affected father and/or brothers). When we performed additional analyses, excluding the six individuals who reported positive family history, the results were similar (not shown).

In the report by Ishii et al. [10], five aberrant *LZTS1* transcripts were described. Our sequence analysis confirmed the presence of an Alu-Sx element 5′ of exon 3. Alu elements are thought to mediate genomic rearrangement [17] and have been implicated in recombination events associated with cancer, as in the cases of Philadelphia chromosome associated with chronic myelogenous leukemia [18] and BRCA1 deletions associated with breast and ovarian cancer [19]. Alu elements have also been known to integrate into exons [20] and into regulatory sites of genes [21]. During the resequencing of *LZTS1*, we found no evidence of missing or rearranged portions of any of the three exons, nor did we find any SNP at splice junctions or splice acceptor sites. Eleven SNP were analyzed in detail within exon 2, intron 2, and exon 3, and four of these SNP had marginally significant association with sporadic prostate cancer cases ($P \leq 0.04$). One SNP flanking the Alu repeat 5′ of exon 3 (WF101–014) was typed in our prostate cancer panels and found to have a marginally significant *P* value (0.04). No direct physical correlation between this SNP and aberrant splicing could be made and no other evidence of deletions or rearrangements were detected. However, because our sequencing strategy involved amplifying and sequencing *LZTS1* using PCR fragments < 600 bp, we can not rule out deletions or rearrangements larger than two overlapping PCR products (~1100 bp). While we cannot draw any conclusions from this evidence, the possibility of Alu-mediated recombination may exist and warrants further study.

This study adds some evidence supporting the role of *LZTS1* as a prostate cancer risk gene. However, this evidence should not be misconstrued to encompass all cancers linked to TSG activity on chromosome 8 without additional studies in specific disease populations. *LZTS1* maps to a very complex region on this chromosome characterized by chromosomal deletions, inversions, and duplications [22]. This region also lacks complete contiguous DNA sequence. The association of *LZTS1* with prostate cancer may be due to the gene's proximity to this aberrant region of chromosome 8. Therefore, association of this region to cancer risk will require an understanding of how the chromosomal aberrations affect the activity of other genes in the same region.

Acknowledgments

We wish to thank Debbie Kiger, Scott Binford, and Catherine Brewer for their technical expertise as well as all of the patients that participated in this study. We would also like to thank Aubrey Turner for critical reading of the manuscript. This work was partially supported by PHS SPORE grants to W.B.I. and J.X.

References

[1] Bova GS, MacGrogan D, Levy A, Pin SS, Bookstein R, Isaacs WB. Physical mapping of chromosome 8p22 markers and their homozygous deletion in metastatic prostate cancer. Genomics 1996;35:46–54.

[2] Cunningham JM, Shan A, Wick MJ, McDonnell SK, Schaid DJ, Tester DJ, Oian J, Takahashi S, Jenkins RB, Bostwick DG, Thibodeau SN. Allelic imbalance and microsatellite instability in prostatic adenocarcinoma. Cancer Res 1996;56:4475–82.

[3] Latil A, Lidereau R. Genetics aspects of prostate cancer. Virchow Arch 1998;432:389–406.

[4] Xu J, Zheng SL, Hawkins GA, Faith D, Kelly B, Isaacs SD, Wiley K, Chang B-L, Ewing C, Bujnovszky P, Carpten J, Bleecker ER, Walsh PC, Trent JM, Meyers DA, Isaacs WB. Linkage and association studies of prostate cancer susceptibility: Evidence for linkage on 8p22–23. Am J Hum Genet 2001;69:341–50.

[5] Gibbs M, Stanford JL, McIndoe RA, Jarvik GP, Kolb S, Goode EL, Chakrabarti L, Schuster EF, Buckley VA, Miller EL, Brandzel S, Li S, Hood L, Ostrander EA. Evidence for a rare prostate cancer-susceptibility locus at chromosome 1p36. Am J Hum Genet 1999;64:776–87.

[6] Goddard KA, Witte JS, Suarez BK, Catalona WJ, Olson JM. Model-free linkage analysis with covariates confirms linkage of prostate cancer to chromosomes 1 and 4. Am J Hum Genet 2001;68:1197–206.

[7] MacGrogan D, Levy A, Bova GA, Isaacs WB, Bookstein R. Structure and methylation-associated silencing of a gene within a homozygously deleted region of human chromosome band 8p22. Genomics 1996;35:55–65.

[8] Bookstein R, Bova GS, MacGrogan D, Levy A, Isaacs WB. Tumor-suppressor genes in prostatic oncogenesis: a positional approach. Br J Urol 1997;79(suppl 1):28–36.

[9] Fujiwara Y, Ohata H, Kuroki T, Koyama K, Tsuchiya E, Monden N, Nakamura Y. Isolation of a candidate tumor suppressor gene on chromosome 8p21.3–p22 that is homologous to an extracellular domain of the *PDGF* receptor beta gene. Oncogene 1995;10:891–5.

[10] Ishii H, Baffa R, Numata SI, Murakumo Y, Rattan S, Inoue H, Mori M, Fidanza V, Alder H, Croce CM. The *FEZ1* gene at chromosome 8p22 encodes a leucine-zipper protein, and its expression is altered in multiple human tumors. Proc Natl Acad Sci U S A 1999;96:3928–33.

[11] Carbeza-Arvelaiz Y, Sepulveda JL, Lebovitz RM, Thompson TC, Chinault AC. Functional identification of *LZTS1* as a candidate prostate tumor suppressor gene on human chromosome 8p22. Oncogene 2001;20:4169–79.

[12] Vecchione A, Ishii H, Shiao Y-H, Trapasso F, Rugge M, Tamburrino JF, Murakumo Y, Alder H, Croce CM, Baffa R. *Fez1/Lzts1* alterations in gastric carcinoma. Clin Cancer Res 2001;7:1546–52.

[13] Ishii H, Vecchione A, Murakumo Y, Baldassarre G, Numata S, Trapasso F, Alder H, Baffa R, Croce CM. *FEZ1/LZTS1* gene at 8p22 suppresses cancer cell growth and regulates mitosis. Proc Natl Acad Sci U S A 2001;98:10374–9.

[14] Weir BS. Genetic data analysis II: Methods for discrete population genetic data. Sunderland, MA: Sinauer Association, Inc. Publishers, 1996.

[15] Zaykin D, Zhivotovsky L, Weir BS. Exact tests for association between alleles at arbitrary numbers of loci. Genetica 1995;96:169–78.

[16] Sasieni PD. From genotypes to genes: doubling the sample size. Biometrics 1999;53:997–1004.

[17] Stenger JE, Lobachev KS, Gordenin D, Darden TA, Jurka J, Resnick MA. Biased distribution of inverted and direct Alus in the human genome: implications for insertion, exclusion, and genome stability. Genom Res 2001;11:12–27.

[18] Kolomietz E, Al-Maghrabi J, Brennan S, Karakova J, Minkin S, Lipton J, Squire JA. Primary chromosomal rearrangements of leukemia are frequently accompanied by extensive submicroscopic deletion and may lead to altered prognosis. Blood 2001;97:3581–8.

[19] Rohlfs EM, Puget N, Graham ML, Weber BL, Garber JE, Skrzynia C, Halperin JL, Lenoir GM, Silverman LM, Mazoyer S. An Alu-mediated 7.1 kb deletion of *BRCA1* exons 8 and 9 in breast and ovarian cancer families that results in alternative splicing of exon 10. Genes Chromosomes Cancer 2000;28:300–7.

[20] Mullersman JE, Pfeffer LM. An Alu cassette in cytoplasmic domain of an interferon receptor subunit. J. Interferon Cytokine Res 1995;15: 815–7.

[21] Swensen J, Hoffman M, Skolnick MH, Neuhausen SL. Identification of a 14 kb deletion involving the promoter region of *BRCA1* in a breast cancer family. Hum Mol Genet 1997;6:1513–7.

[22] Giglio S, Roman KW, Matsumoto N, Calvari V, Gimelli G, Neuman T, Ohashi H, Voullaire L, Larizza D, Giorda R, Weber JL, Ledbetter DH, Zuffardi O. Olfactory receptor-gene clusters, genomic-inversions polymorphisms and common chromosome rearrangements. Am J Hum Genet 2001;68:874–83.

Int. J. Cancer: **98,** 938–942 (2002)

DOI 10.1002/ijc.10290

Publication of the International Union Against Cancer

SHORT REPORT

TRANSMISSION/DISEQUILIBRIUM TESTS OF ANDROGEN RECEPTOR AND GLUTATHIONE S-TRANSFERASE PI VARIANTS IN PROSTATE CANCER FAMILIES

Gloria Y.F. Ho[1]*, Michael Knapp[2], Diha Freije[3], William G. Nelson[4], Jeffrey R. Smith[5], John D. Carpten[5], Joan E. Bailey-Wilson[5], Terri H. Beaty[6], Gloria Petersen[7], Jianfeng Xu[8], Victor Kamensky[1], Patrick C. Walsh[3] and William B. Isaacs[3]

[1]*Department of Epidemiology and Social Medicine, Albert Einstein College of Medicine, Bronx, NY, USA*
[2]*Institute for Medical Biometry, Informatics and Epidemiology, University of Bonn, Bonn, Germany*
[3]*Department of Urology, Johns Hopkins Medical Institutions, Baltimore, MD, USA*
[4]*Oncology Center, Johns Hopkins Medical Institutions, Baltimore, MD, USA*
[5]*National Human Genome Research Institute, National Institute of Health, Bethesda, MD, USA*
[6]*Department of Epidemiology, Johns Hopkins School of Hygiene and Public Health, Baltimore, MD, USA*
[7]*Mayo Clinic, Rochester, MN, USA*
[8]*Center for Genomics, Wake Forest University School of Medicine, Winston-Salem, NC, USA*

Population-based case-control studies have found relationships between risk of prostate cancer and genetic polymorphisms in the CAG repeat and GGC repeat of the X-linked androgen receptor gene *(AR)* as well as the autosomal gene coding for glutathione S-transferase pi *(GSTP1)*. This family-based study utilized the transmission disequilibrium test to examine whether there was evidence that these polymorphisms could account for familial aggregation of prostate cancer. Seventy-nine North American pedigrees were studied. Most of these families had 3 or more affected first-degree relatives. Genotype information was obtained on 578 individuals. The reconstruction combined transmission disequilibrium test (RC-TDT) was used to test for linkage. There was no evidence of linkage to the CAG and GGC repeat sequences in the *AR* gene or the pentanucleotide (ATAAA) repeat in the *GSTP1* gene when each allele was analyzed separately or when alleles were grouped by repeat length. Our findings do not support the hypothesis that familial clustering of prostate cancer in high-risk families is attributable to these genetic variants.

Key words: *prostate cancer; androgen receptor; glutathione S-transferase pi; transmission disequilibrium test*

Population-based case-control studies have found associations between prostate cancer and polymorphisms of the androgen receptor gene *(AR)* and glutathione S-transferase pi gene *(GSTP1)*.[1–8] False-positive results could arise from population-based association studies due to population stratification. Such bias occurs when the study population is ethnically heterogeneous and 1 or more ethnic subgroups have both a higher prevalence of an allele and a higher risk of prostate cancer; it would then appear that the allele is related to prostate cancer.[9,10] The demonstration of genetic linkage of risk to prostate cancer to the *AR* and *GSTP1* genes in family-based studies, which are not affected by population subdivision and admixture, would strengthen the argument for their role in the etiology of prostate cancer.[11] Our study in multiplex prostate cancer families examined whether genetic variations in the *AR* and *GSTP1* genes could account for familial aggregation of prostate cancer in high-risk families.

The androgen receptor is an intracellular receptor that binds to testosterone and dihydrotestosterone and induces transcription of androgen-responsive genes in target cells. Since it plays a direct role in the growth of prostate cells, the *AR* gene on the X chromosome (Xq11-12) is hypothesized to be a susceptibility gene for prostate cancer. The amino-terminal domain of the AR protein, which is important for transcriptional activity, is encoded by exon 1.[12] This exon contains 2 trinucleotide repeat sequences: CAG (glutamine) and GGC (glycine). The length of the CAG repeat appears to be inversely correlated with the transactivation function of AR,[13] suggesting shorter CAG repeat may cause more rapid growth of prostate cells.[14] Several case-control studies did find increased risk of prostate cancer to be associated with shorter CAG and GGC repeat sequences.[1–6]

The glutathione S-transferase supergene family plays a central role in the detoxification of several potential carcinogens. The π subfamily *(GSTP1)* on chromosome 11 (11q13) is involved in the inactivation of carcinogens in cigarette smoke.[7] *GSTP1* is considered a potential candidate gene for prostate cancer, since nearly all prostate cancer tissue displays hypermethylation at the *GSTP1* promoter region, which is accompanied by a marked decrease in gene expression.[15] In our study, we examined polymorphism of the pentanucleotide repeat (ATAAA) in the 5′ promoter region of *GSTP1*.[16] This distinct sequence defines the boundary of the methylated and unmethylated regions in the *GSTP1* promoter in normal tissues and may act as a barrier to the methylation of *GSTP1*.[17]

MATERIAL AND METHODS

Methods of obtaining 79 North American pedigrees were published previously.[18] Briefly, 65% of the families were identified by referrals from urologists throughout the country, 23% were identified by family history records of the patient population seen at the Johns Hopkins Hospital for treatment of prostate cancer and the remainder of the families responded to lay publications describing the ongoing genetic project on prostate cancer.[19] A family was considered to have hereditary prostate cancer and eligible to participate if it fulfilled 1 of 3 criteria: *(i)* prostate cancer occurred in ≥3 first-degree relatives; *(ii)* ≥2 men had prostate cancer diagnosed before age 55; or *(iii)* prostate cancer occurred in ≥3 successive generations. Prostate cancer diagnosis was confirmed by medical records, pathologic reports, or both. Table I describes the characteristics of these 79 families, of which 97% were Caucasians. Genotype information and affection status were available

*Correspondence to: Department of Epidemiology and Social Medicine, Albert Einstein College of Medicine, Belfer Building 1312, Bronx, NY 10461, USA. Fax: +718-430-3076. E-mail: ho@aecom.yu.edu

Received 14 August 2001; Revised 19 November 2001; Accepted 4 December 2001

Published online 20 February 2002

TABLE I – CHARACTERISTICS OF PROSTATE CANCER FAMILIES, AMONG INDIVIDUALS WITH KNOWN GENOTYPE AND AFFECTION STATUS[1]

Characteristic	No.
Families	79
Subjects	711
Subjects with known genotypes and affection status	578
Typed:reconstructed	476 (82%):102 (18%)
Male:female	384 (66%):194 (34%)
Affected men:unaffected men	318 (83%):66 (17%)
Affected men per family (average)	4.0 (range 2–13)
Average age at diagnosis	65.2 (range 39–86)
Average age at blood drawn	
Affected men	68.2 (range 40–89)
Unaffected men	64.3 (range 35–89)

[1]Among individuals with genotype information on at least 1 of the 3 loci (*AR* CAG repeat, *AR* GGC repeat or *GSTP1*).

for 578 (81%) of the subjects. Of these, genotypes of 102 (18%) were unequivocally inferred from other genotyped family members. Genotypes were imputed for the purpose of the reconstruction combined transmission disequilibrium test described below.

DNA typing methods

DNA was purified from lymphoblastoid cell lines using a phenol/chloroform extraction protocol followed by ethanol precipitation. In a previous genome-wide scan study, we extensively compared genotypes from blood DNA and lymphoblastoid DNA from a sample of subjects and there was complete agreement of genotypes obtained with the 2 sources of DNA.

For the *AR* gene, ^{32}P-labeled oligonucleotide primers that flank the CAG repeat (5′-CTTTCCAGAATCTGTTCCAGAGC-3′ and 5′-GCTGTGAAGGTTGCTGTTCCTC-3′) and GGC repeat (5′-TGGCACACTCTCTTCACAGCCGA-3′ and 5′-GTACCACACATCAGGTGCGGTG-3′) were used to generate PCR products from genomic DNA. The products were analyzed on 6% denaturing polyacrylamide gels and the repeat lengths were determined by comparison with products of known repeat length as determined by sequencing. A polymorphic imperfect pentanucleotide (ATAAA) repeat in the *GSTP1* promoter was genotyped as described by Harada *et al.*,[16] with the exception that the PCR products were analyzed on 6% denaturing polyacrylamide gels. Genotyping of the CAG repeat was carried out in duplicate in 2 separate laboratories and there were no discrepancies. For a subset of subjects, direct sequencing of PCR products was carried out for both the *AR* repeat polymorphisms and for the *GSTP1* polymorphism. In all cases, the sequence data were consistent with the allele identification as scored by genotyping.

Statistical analysis

The transmission/disequilibrium test (TDT) detects linkage between marker and disease loci in the presence of linkage disequilibrium.[9] It requires families in which both parents and at least 1 affected offspring are genotyped. When parental genotypes are missing, it may be possible to reconstruct them from the genotypes of their affected and, if available, unaffected offspring. Such a reconstruction procedure, however, can introduce bias.[20] The sib-TDT, on the other hand, needs genotypes of at least 1 affected offspring and 1 unaffected sibling, rather than genotypes of the parents.[21] The reconstruction combined TDT (RC-TDT), introduced by Knapp, allows parental-genotype reconstruction in the TDT, corrects for the biases resulting from such reconstruction and combines data from the TDT and the sib-TDT. Hence, the RC-TDT utilizes information from families in which parental genotypes are either typed or reconstructed as well as families in which parental genotypes are not available but genotypes of unaffected sibs are available.[20]

In the allele-by-allele RC-TDT, each allele of the *AR* and *GSTP1* genes, designated as allele M, was tested separately against all other alleles grouped together. For the *GSTP1* analysis, families were classified into 1 of 4 possible categories as defined by Knapp:[20] *(i)* both parents were genotyped and at least 1 parent was heterozygous for allele M; *(ii)* only 1 parent was typed, the genotype of the missing parent was reconstructed and at least 1 parent was heterozygous for allele M; *(iii)* both parental genotypes were reconstructed and at least 1 parent was heterozygous for allele M; or *(iv)* genotypes of both parents were not available, but conditions for the sib-TDT were fulfilled. All families not belonging to categories 1–4 were excluded from this analysis. For families in category 1, the expected number of transmissions of allele M from the heterozygous parents to the affected offspring and its variance were calculated based on the methods for TDT.[9] For families in category 4, the expectation and variance of the number of M alleles in affected offspring under the null hypothesis of no linkage were computed using the equations for sib-TDT.[21] For families in categories 2 and 3, formulas for the reconstructed TDT, which consider genotype reconstruction, were used.[20] The observed and expected numbers of M alleles were then combined across all families in the test statistics of the RC-TDT. Because of the late onset age of prostate cancer, the true affection status of the presently unaffected offspring might be problematic in the sib-TDT. Hence, families were classified preferentially into category 1, 2 or 3 whenever possible even though they might have also fulfilled requirements for the sib-TDT.

The test statistics of TDT, sib-TDT and RC-TDT published to date apply only to autosomes. These methods have recently been extended to test for linkage between X-linked markers and diseases that affect either males only or both sexes. To perform RC-TDT for the X-linked *AR* gene when only males could have the disease, families were classified into 1 of 3 possible categories: *(i)* the mother was genotyped and heterozygous for allele M; *(ii)* the genotype of the mother was reconstructed and she was heterozygous for allele M; or *(iii)* the genotype of the mother was not available, but the conditions for sib-TDT were fulfilled. All families that could not be classified into 1 of these 3 categories were excluded from the analysis. The modified equations for TDT, reconstructed TDT and sib-TDT for X-linked markers and sex-limited diseases were used for families in categories 1, 2 and 3, respectively, to calculate the expected number of allele M and its variance.[22,23] The observed and expected numbers of M alleles were summed across all families in the test statistics of X-linked RC-TDT.

The results of the X-linked and autosomal RC-TDTs are shown in Tables II–IV and some definitions are explained here. Basically, the null hypothesis was that prostate cancer and the marker tested were unlinked—there was no excess transmission of allele M from heterozygous parents to their affected offspring (for the TDT), or there was no difference in the frequency of allele M between the affected and unaffected offspring (for the sib-TDT). The test statistics of RC-TDT examined whether the observed number of M alleles among the affected individuals (column 5) was different from expected (column 6). Two-sided exact *p*-values were computed (column 7).[24] The total number of families analyzed in the allele-by-allele analysis of a particular polymorphism (column 4) could add up to more than or less than 79. A family could be analyzed more than once if it carried multiple alleles of the polymorphism, or a family might not be analyzed at all if it did not fit into 1 of the eligibility categories for the RC-TDT described above. Also presented in Tables II–IV are allele frequencies among the affected and unaffected male offspring in all families (columns 2–3), including families that did not fit the eligibility categories for the RC-TDT. These allele frequencies are shown for descriptive purposes only; they were computed from dependent observations within families and the test statistics of the RC-TDT was not based on these frequencies.

The RC-TDT is a test of linkage between marker and disease. It is also a test of association only if the data are entirely from simplex families—1 affected offspring per family for the TDT and

TABLE II – ALLELE FREQUENCIES AND THE X-LINKED RC-TDT FOR THE ANDROGEN RECEPTOR CAG REPEAT IN PROSTATE CANCER[1]

Allele M (expressed as no. of repeats)	Allele frequency (%) in male subjects		X-linked RC-TDT in families			
	Affected offspring (Total no. of alleles = 273)	Unaffected offspring (Total no. of alleles = 52)	No. of families[2]	Observed no. of M alleles among affected	Expected no. of M alleles among affected	*p*-value
Individual alleles						
14	0.73	1.92	1	2	2.0	1.0
15	0.37	1.92	1	1	0.5	1.0
16	1.47	3.85	2	2	3.4	0.433
17	2.56	1.92	4	7	7.1	1.0
18	6.96	1.92	5	7	6.6	1.0
19	4.03	5.77	5	6	7.2	0.682
20	10.99	19.23	20	27	32.6	0.148
21	12.09	11.54	15	26	26.2	1.0
22	11.72	13.46	13	16	17.6	0.671
23	10.62	25.00	17	26	25.7	1.0
24	13.19	0	16	30	24.7	0.110
25	10.99	0	12	21	17.4	0.196
26	7.69	7.69	9	13	14.6	0.663
27	1.47	1.92	2	2	3.5	0.393
28	2.56	0	4	7	6.5	1.0
29	2.56	3.85	5	7	6.0	0.763
Alleles grouped by length of repeats						
< 22	39.19	48.08	33	41	48.0	0.136
22–23	22.34	38.46	20	34	34.6	0.985
24–25	24.17	0	23	43	35.3	0.039
≥ 26	14.29	13.46	16	22	23.4	0.779

[1]Allele M refers to the particular allele under analysis.–[2]All families were in categories 1 and 2, where either X-linked TDT or reconstructed TDT was applied (see definitions for eligibility categories in Statistical Analyses).

TABLE III – ALLELE FREQUENCIES AND X-LINKED RC-TDT FOR ANDROGEN RECEPTOR GGC REPEAT IN PROSTATE CANCER[1]

Allele M (expressed as no. of repeats)	Allele frequency (%) in male subjects		X-linked RC-TDT in families			
	Affected offspring (Total no. of alleles = 277)	Unaffected offspring (Total no. of alleles = 55)	No. of families[2]	Observed no. of M alleles among affected	Expected no. of M alleles among affected	*p*-value
Individual alleles						
10	0.72	5.45	2	2	2.0	1.0
12	0.72	1.82	3	2	5.4	0.045
14	2.17	0	2	6	5.0	0.754
15	0.36	5.45	1	1	1	1.0
16	63.18	58.18	31	60	52.6	0.129
17	27.80	27.27	24	38	43.0	0.272
18	2.89	0	4	8	7.0	0.723
20	1.08	1.82	2	1	3	0.219
21	1.08	0	1	3	2	0.571
Alleles grouped by length of repeats						
≤ 16	67.15	70.91	29	56	51.0	0.310
> 16	32.85	29.09	29	46	51.0	0.310

[1]Allele M refers to the particular allele under analysis.–[2]All families were in categories 1 and 2, where either X-linked TDT or reconstructed TDT was applied (see definitions for eligibility categories in Statistical Analyses).

exactly 1 affected and 1 unaffected sib in each family for the sib-TDT. In our study, some families consisted of multiple affected and unaffected offspring and hence the RC-TDT was a test of linkage but not a valid test of association.

RESULTS

The number of CAG repeats in the *AR* gene ranged from 14–29. None of the 16 alleles, when analyzed individually by the X-linked RC-TDT, showed linkage with prostate cancer (Table II). The alleles were then grouped by length of repeats using categorization from previous publications in order to allow comparison of results across studies.[1,3,25] Although the allele with 24-25 CAG repeats was more likely to occur among the affected individuals than expected ($p = 0.039$), this difference was not significant when adjusted for multiple comparisons. There was also no apparent trend of excess transmission of the alleles with short CAG repeat length to affected individuals (Table II). Nine alleles, with the number of repeats ranging from 10–21, were detected for the GGC repeat in the *AR* gene. However, the alleles with 16 and 17 repeats predominated, while the other alleles occurred in no more than 4 families. There was no evidence of linkage when alleles were analyzed individually or grouped into 2 categories as in previous studies,[3,6] ≤16 and >16 repeats (Table III). Twenty-seven alleles were detected for the *GSTP1* gene. Presented in Table IV are the RC-TDT results of 7 common alleles and a combined group of 20 rare alleles, in which the frequency of each allele was less than 1%. There was no evidence of linkage to the *GSTP1* gene.

DISCUSSION

The length of the CAG repeat in the *AR* gene of humans varies from 11–33 repeats, with a modal length of about 20. Because of the inverse correlation between length of the CAG repeat and transactivation activity of *AR*,[13] several studies have examined the association between length of the CAG repeat and the natural history of prostate cancer. A case-control study in Caucasian men showed a 3% decrease in risk of prostate cancer for each CAG

TABLE IV – ALLELE FREQUENCIES AND THE RC-TDT FOR THE PENTANUCLEOTIDE (ATAAA) REPEAT OF *GSTP1* IN PROSTATE CANCER[1]

Allele M (in bp)	Allele frequency (%) in male subjects		RC-TDT in families			
	Affected offspring (Total no. of alleles = 558)	Unaffected offspring (Total no. of alleles = 108)	No. of families (TDT, sib-TDT)[2]	Observed no. of M alleles among affected	Expected no. of M alleles among affected	*p*-value
187	43.37	53.70	40 (26, 14)	112	111.8	1.0
195	2.33	1.85	10 (10, 0)	12	12.7	0.950
199	2.69	0.93	1 (1, 0)	9	6.2	0.155
201	9.86	8.33	17 (16, 1)	26	26.0	1.0
204	1.79	1.85	5 (4, 1)	8	7.6	1.0
207	25.63	21.30	33 (25, 8)	54	60.5	0.208
213	2.15	0.93	6 (4, 2)	7	5.8	0.654
Others[3]	12.18	11.11	13 (8, 5)	27	26.3	0.950

[1]Allele M refers to the particular allele under analysis. The alleles were named based on the electrophoresis patterns of the PCR products. The size of each amplified DNA fragment (in base pairs) was determined by comparison with products of known repeat length as determined by sequencing.–[2]Parentheses contain the number of families in categories 1–3, where either TDT or reconstructed TDT was applied, and the number of families in category 4, where sib-TDT was used (see definitions for eligibility categories in Statistical Analyses).–[3]The following 20 alleles, each with an allele frequency of ≤ 1%, were combined: 166, 169, 173, 181, 190, 192, 196, 197, 198, 200, 202, 203, 205, 208, 209, 210, 211, 216, 219, 225.

repeat and an odds ratio (OR) of 1.23 when men with <22 repeats were compared to those with ≥22 repeats.[3] A nested case-control study conducted in the Physician's Health Study cohort, which is predominantly Caucasian, reported significant associations between a shorter CAG repeat sequence and risk of total prostate cancer (OR = 1.52) as well as advanced cancer with Gleason grade ≥7 or stage C or D at diagnosis (OR = 2.14).[5] Other studies also found length of CAG repeat to be related to the aggressiveness of prostate cancer or early age of diagnosis.[2,4]

The GGC repeat in *AR* is 1.1 kb from the CAG repeat, but it does not appear to be critical for *AR* transactivation.[12] A few studies have examined polymorphism of the GGN repeat, which is the GGC repeat together with the preceding $(GGT)_3GGG(GGT)_2$ sequence. Although 1 case-control study showed the risk of prostate cancer to be higher in men with ≤16 GGN repeats relative to those with >16 repeats (OR = 1.60), another study only found a modest association (OR = 1.20).[3,6]

GSTP1 is suspected to play a role in prostate cancer, since almost all cancer tissue specimens show hypermethylation at the *GSTP1* promoter region and loss of expression of this enzyme.[15] Some studies found association between prostate cancer and a single nucleotide polymorphism at base pair 313 of the *GSTP1* gene.[7,8] The pentanucleotide repeat (ATAAA) polymorphism of *GSTP1* gene reported in our article has not been investigated in other epidemiologic studies of prostate cancer. This repeat sequence in the 5′ promoter region of *GSTP1* may act as a barrier to the methylation of *GSTP1*.[17]

Previous studies suggest that genetic polymorphisms at the CAG and GGC repeats in the *AR* gene and in the *GSTP1* gene have a modest association with prostate cancer. There are several interpretations. First, the *AR* and/or *GSTP1* gene could be a susceptibility gene and is accountable for a proportion of prostate cancers. Second, the *AR* and/or *GSTP1* gene could be linked to a susceptibility locus, which causes some prostate cancers and there is disequilibrium between the 2 loci. Third, the associations could simply be due to population stratification or confounding in these case-control studies. Direct evidence for linkage in prostate cancer families would therefore lend support to the first and second suppositions and strengthen the roles of these genes as susceptibility genes or genetic markers for prostate cancer.

One affected sib-pair study with 41 sib-pairs and 6 sib-trios of Caucasian men did not find a higher than expected concordance rate at the CAG repeat locus.[26] RC-TDT, a statistical method for linkage, was applied to prostate cancer families in our study. There was no evidence for linkage to the CAG and GGC repeats in the *AR* gene or to the pentanucleotide repeat in the *GSTP1* gene. Of note is that parametric and nonparametric linkage analyses were also performed and did not show evidence for linkage (data not shown). The null finding could be due to the small sample size and lack of statistical power, but our data did not even show a trend in relationship between shorter CAG repeat and risk of prostate cancer as reported in previous studies. Nevertheless, our findings do not necessarily negate an etiologic role for these genes. The families in our study were selected to show substantial familial clustering with ≥3 first-degree affected relatives, early age of diagnosis in ≥2 men or presence of prostate cancer in ≥3 successive generations. Mutations in the major susceptibility genes that cause familial aggregation and have a relatively high penetrance would be overrepresented in these families. Other susceptibility genes, particularly those that are associated with sporadic cases or not typically found in multiplex families, may be unlikely to show linkage in these selected high-risk families. In conclusion, our findings do not support the hypothesis that familial clustering of prostate cancer in high-risk families is attributable to genetic variants in the CAG and GGC repeats in the *AR* gene or in the pentanucleotide repeat in the *GSTP1* gene.

ACKNOWLEDGEMENTS

Supported in part by Public Health Service grants R01 CA64247 (G.Y.F.H.) and SPORE CA58236 (W.B.I.) from the National Cancer Institute, National Institutes of Health, Department of Health and Human Services as well as DAMD 17-98-1-8469 (W.B.I.) from the Department of Defense.

REFERENCES

1. Irvine RA, Yu MC, Ross RK, et al. The CAG and GGC microsatellites of the androgen receptor gene are in linkage disequilibrium in men with prostate cancer. Cancer Res 1995;55:1937–40.
2. Hardy DO, Scher HI, Bogenreider T, et al. Androgen receptor CAG repeat lengths in prostate cancer: correlation with age of onset. J Clin Endocrinol Metab 1996;81:4400–5.
3. Stanford JL, Just JJ, Gibbs K, et al. Polymorphic repeats in the androgen receptor gene: molecular markers of prostate cancer risk. Cancer Res 1997;57:1194–8.
4. Ingles S, Ross RK, Yu MC, et al. Association of prostate cancer risk with genetic vitamin D receptor and androgen receptor. J Natl Cancer Inst 1997;89:166–70.
5. Giovannucci E, Stampfer MJ, Krithivas K, et al. The CAG repeat within the androgen receptor gene and its relationship to prostate cancer. Proc Natl Acad Sci USA 1997;94:3320–3.
6. Platz EA, Giovannucci E, Dahl D, et al. The androgen receptor gene GGN microsatellite and prostate cancer risk. Cancer Epidemiol Biomark Prev 1998;7:379–84.
7. Harries LW, Stubbins MJ, Forman D, et al. Identification of genetic polymorphisms at the glutathione S-transferase Pi locus and associa-

tion with susceptibility to bladder, testicular and prostate cancer. Carcinogenesis 1997;18:641–4.
8. Gsur A, Haidinger G, Hinteregger S, et al. Polymorphisms of glutathione-S-transferase genes *(GSTP1, GSTM1* and *GSTT1)* and prostate-cancer risk. Int J Cancer 2001;95:152–5.
9. Spielman RS, McGinnis RE, Ewens WJ. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). Am J Hum Genet 1993;52:506–16.
10. Wacholder S, Rothman N, Caporaso N. Population stratification in epidemiologic studies of common genetic variants and cancer: quantification of bias. J Natl Cancer Inst 2000;92:1151–8.
11. Ewens WJ, Spielman RS. The transmission/disequilibrium test: history, subdivision and admixture. Am J Hum Genet 1995;57:455–64.
12. Choong CS, Wilson EM. Trinucleotide repeats in the human androgen receptor: a molecular basis for disease. J Mol Endocrinol 1998;21: 235–57.
13. Kazemi-Esfarjani P, Trifiro MA, Pinsky L. Evidence for a repressive function of the long polyglutamine tract in the human androgen receptor: possible pathogenetic relevance for the (CAG)n-expanded neuronopathies. Hum Mol Genet 1995;4:523–7.
14. Schoenberg MP, Hakimi JM, Wang S, et al. Microsatellite mutation in the androgen receptor gene in human prostate cancer. Biochem Biophys Res Comm 1994;198:74–80.
15. Lee W-H, Isaacs WB, Bova GS, et al. CG island methylation changes near the GSTP1 gene in prostatic carcinoma cells detected using the polymerase chain reaction: a new prostate cancer biomarker. Cancer Epidemiol Biomark Prev 1997;6:443–50.
16. Harada S, Nakamura T, Misawa S. Polymorphism of pentanucleotide repeats in the 5′ flanking region of glutathione S-transferase (GST) π gene. Hum Genet 1994;93:223–4.
17. Millar DS, Paul CL, Molloy PL, et al. A distinct sequence $(ATAAA)_n$ separates methylated and unmethylated domains at the 5′-end of the *GSTP1* CpG island. J Biol Chem 2000;275:24893–9.
18. Grönberg H, Isaacs SD, Smith JR, et al. Characteristics of prostate cancer in families potentially linked to the hereditary prostate cancer 1 (HPC1) locus. J Am Med Assoc 1997;278:1251–6.
19. Smith JR, Freije D, Carpten JD, et al. Major susceptibility locus for prostate cancer on chromosome 1 suggested by a genome-wide search. Science 1996;274:1371–4.
20. Knapp M. The transmission/disequilibrium test and parental-genotype reconstruction: the reconstruction-combined transmission/disequilibrium test. Am J Hum Genet 1999;64:861–70.
21. Spielman RS, Ewens WJ. A sibship test for linkage in the presence of association: the sib transmission/disequilibrium test. Am J Hum Genet 1998;62:450–8.
22. Ho GYF, Bailey-Wilson J. The transmission/disequilibrium test for linkage on the X chromosome. Am J Hum Genet 2000;66:1158–60.
23. Horvath S, Laird NM, Knapp M. The transmission/disequilibrium test and parental-genotype reconstruction for X-chromosomal markers. Am J Hum Genet 2000;66:1161–7.
24. Knapp M. Using exact p values to compare the power between the reconstruction-combined transmission/disequilibrium test and the sib transmission/disequilibrium test. Am J Hum Genet 1999;65:1208–10.
25. Lange EM, Chen H, Brierley K, et al. The polymorphic exon 1 androgen receptor CAG repeat in men with a potential inherited predisposition to prostate cancer. Cancer Epidemiol Biomark Prev 2000;9:439–42.
26. Sun S, Narod SA, Aprikian A, et al. Androgen receptor and familial prostate cancer. Nat Med 1995;1:848–9.

Reprint Series
22 November 1996, Volume 274, pp. 1371–1374

SCIENCE

Major Susceptibility Locus for Prostate Cancer on Chromosome 1 Suggested by a Genome-Wide Search

Jeffrey R. Smith,* Diha Freije,* John D. Carpten,*
Henrik Grönberg,* Jianfeng Xu,* Sarah D. Isaacs,
Michael J. Brownstein, G. Steven Bova, Hong Guo,
Piroska Bujnovszky, Deborah R. Nusskern, Jan-Erik Damber,
Anders Bergh, Monika Emanuelsson, Olli P. Kallioniemi,
Jennifer Walker-Daniels, Joan E. Bailey-Wilson, Terri H. Beaty,
Deborah A. Meyers, Patrick C. Walsh, Francis S. Collins,
Jeffrey M. Trent,† and William B. Isaacs

Major Susceptibility Locus for Prostate Cancer on Chromosome 1 Suggested by a Genome-Wide Search

Jeffrey R. Smith,* Diha Freije,* John D. Carpten,* Henrik Grönberg,* Jianfeng Xu,* Sarah D. Isaacs, Michael J. Brownstein, G. Steven Bova, Hong Guo, Piroska Bujnovszky, Deborah R. Nusskern, Jan-Erik Damber, Anders Bergh, Monika Emanuelsson, Olli P. Kallioniemi, Jennifer Walker-Daniels, Joan E. Bailey-Wilson, Terri H. Beaty, Deborah A. Meyers, Patrick C. Walsh, Francis S. Collins, Jeffrey M. Trent,† William B. Isaacs

Despite its high prevalence, very little is known regarding genetic predisposition to prostate cancer. A genome-wide scan performed in 66 high-risk prostate cancer families has provided evidence of linkage to the long arm of chromosome 1 (1q24-25). Analysis of an additional set of 25 North American and Swedish families with markers in this region resulted in significant evidence of linkage in the combined set of 91 families. The data provide strong evidence of a major prostate cancer susceptibility locus on chromosome 1.

Prostate cancer is the most common malignancy diagnosed in U.S. males, accounting for more than 40,000 deaths in this country annually (*1*). African Americans have the highest incidence and mortality rates of any population studied (*2*). Numerous studies have provided evidence for familial clustering of prostate cancer, indicating that family history is a major risk factor for this disease (*3–5*). Segregation analysis of familial prostate cancer suggests the existence of at least one dominant susceptibility locus and predicts that rare high-risk alleles at such loci account in the aggregate for 9% of all prostate cancers and more than 40% of early onset disease (*6*).

Analyses of genetic alterations in prostate cancer have demonstrated frequent duplication of DNA sequences on the distal long arm of chromosome 8 (*7*), as well as loss of DNA sequences resulting in loss of heterozygosity (LOH) for the short arm of chromosome 8 and the long arm of chromosome 13 (*8, 9*). Preliminary investigations by linkage analysis of these regions as well as sites of known tumor suppressor genes have not identified a susceptibility locus in prostate cancer (*10, 11*).

Prostate cancer presents a number of serious obstacles to linkage analysis. The prevalence is extremely high; there is a one in five lifetime probability of prostate cancer diagnosis in U.S. males (*1*). This potentially could

result in a high rate of phenocopies; individuals whose prostate cancers result from very different causes. The late age of onset [less than 0.1% of prostate cancer cases are diagnosed under the age of 40 (*1*)] leads to general lack of available samples from an affected individual's ancestors. These obstacles are complicated by the absence of known clinical features (other than age of onset) that might allow subgrouping of prostate cancer families to reflect potential genetic heterogeneity (*5*). Finally, it is difficult to find extended pedigrees that are highly informative for linkage (in other words, that contain large numbers of affected family members) (*12*).

In spite of these difficulties, we have undertaken a linkage analysis to search for evidence of loci contributing to risk for prostate cancer in a group of 79 North American and 12 Swedish pedigrees, each having at least three first-degree relatives affected with prostate cancer. These families were selected on the basis of the number of affected males from which samples could be obtained for typing, either as blood samples or archival specimens and the absence of evidence of bilineal inheritance (*13*). A summary of the characteristics of the families studied is given in Table 1. Overall, affected individuals in these families had an average age of diagnosis of 65, with a total of 34 males diagnosed before the age of 55.

To search for the location of high-risk alleles for prostate cancer, a genome-wide scan was performed in a subgroup of 66 North American families. A total of 341 dinucleotide repeat markers were analyzed in these pedigrees to complete a map with a marker density of 10 cM (*14*), requiring more than 130,000 genotypes. On average, 79% of our study group were heterozygous for each marker. For the parametric analysis of the genotype data, we used a model of dominant inheritance that includes a fixed phenocopy rate of 15% and the assumption that unaffected men over the age of 75 are not likely to be gene carriers (*15*). A plot of two-point lod (logarithm of the likelihood ratio for linkage) scores (*16*) for the genome-wide scan ($\hat{Z}$) is shown in Fig. 1. The highest lod score observed was 2.75 with marker *D1S218*, which maps to the distal long arm of chromosome 1 (1q24-25). As chromosome 1 showed the most significant evidence for linkage, additional markers in this region were typed in the original 66 families as well as in an additional group of 25 families, 12 of which were collected in Sweden (*13*). These analyses provided additional evidence for linkage in the 1q24-25 region with a maximum two-point lod of 3.65 at recombination fraction $\Theta = 0.18$ with marker *D1S2883* (Table 2).

As parametric analyses are model-dependent, we also used nonparametric analyses to further examine linkage data in this region (*16*). Nonparametric multipoint linkage (NPL) Z scores are given for this analysis in Table 2. Highly significant *P*-values were obtained for multiple markers, providing further evidence for linkage in this region. To determine the most likely location for the susceptibility locus, parametric multipoint analyses were performed with various combinations of markers in this region. Lod scores >4 were obtained, but did not allow unequivocal placement of the susceptibility locus due to apparent genetic heterogeneity. Significant evidence for locus heterogeneity ($\chi^2 = 8.11$, $P = 0.004$) (*16*) was obtained by an admixture test with an estimate of 34% of the families being linked to the region. The maximum multipoint lod score with markers *D1S2883*, *D1S158*, and *D1S422* under the assumption of heterogeneity was 5.43, with the postulated susceptibility locus mapping close to *D1S422* (Fig. 2). No clinical features appeared to distinguish families showing linkage to chromosome 1 from the non-linked pedigrees.

The risk of prostate cancer in siblings of affected individuals is modified by the age of diagnosis (*6*). Subgrouping families by age of diagnosis, either by mean age within a family or by number of men diagnosed under age 55, provided little evidence that the families showing linkage to chromosome 1 had an earlier onset of prostate cancer than the unlinked families. However, because of difficulties in equating age of diagnosis with age of onset (*17*), further analysis will be necessary to support this conclusion.

Both African-American families analyzed in this study showed linkage to this region, yielding a combined lod score of 1.4. As there is evidence of linkage in Caucasian families in Sweden and North America as well, alterations in the 1q24-25 region may increase prostate cancer susceptibility in a variety of populations and ethnic backgrounds.

LOH studies have not previously implicated the chromosomal region 1q24-25 in

Table 1. Prostate cancer families.

Sample	Number of families	Average number per family (range)		Average age of diagnosis (range)
		Affected	Typed*	
North American	79	5.1 (3–15)	3.7 (2–11)	64.3 (39–85)
Swedish	12	3.9 (3–5)	3.6 (3–5)	69.3 (56–76)
Total	91	4.9 (3–15)	3.7 (2–11)	64.9 (39–85)

*Typed refers to the number of affected family members analyzed.

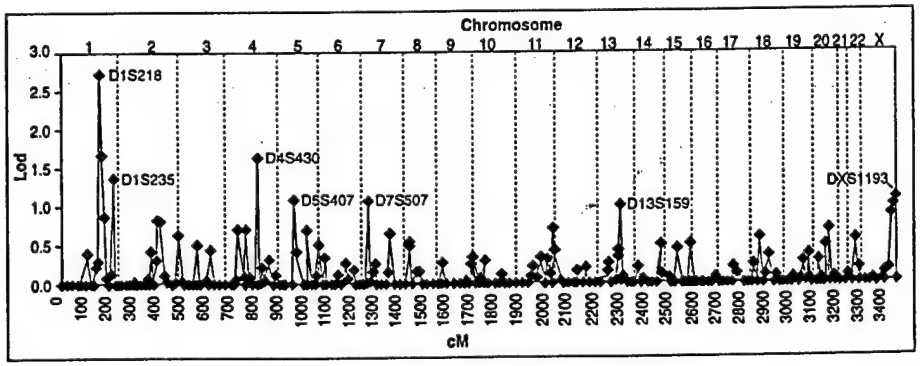


Fig. 1. Two-point lod scores for the genome-wide scan. Affected and unaffected individuals in 66 prostate cancer pedigrees were genotyped at 341 loci throughout the genome. Maximum two-point lod scores were calculated and the results plotted as a function of marker location in centimorgans. Chromosomal number is designated at the top of the plot.

J. R. Smith, J. D. Carpten, O. P. Kallioniemi, J. Walker-Daniels, J. E. Bailey-Wilson, F. S. Collins, J. M. Trent, National Center for Human Genome Research, National Institutes of Health, Bethesda, MD.
D. Freije, H. Grönberg, S. D. Isaacs, G. S. Bova, H. Guo, P. Bujnovszky, D. R. Nusskern, P. C. Walsh, W. B. Isaacs, James Buchanan Brady Urologic Institute; J. Xu and D. A. Meyers, Department of Medicine, Johns Hopkins University, School of Medicine, Baltimore, MD.
M. J. Brownstein, National Institute of Mental Health, National Institutes of Health, Bethesda, MD.
T. H. Beaty, Department of Epidemiology, Johns Hopkins University, School of Hygiene and Public Health, Baltimore, MD.
J. R. Smith, Department of Internal Medicine, University of Michigan, Ann Arbor, MI.
H. Grönberg and M. Emanuelsson, Department of Oncology; J.-E. Damber, Department of Urology; A. Bergh, Department of Pathology, Umeå University, Umeå, Sweden.

*These authors contributed equally to this study.
†To whom correspondence should be addressed. E-mail: jtrent@nchgr.nih.gov

prostate cancer, although analysis of cancer DNA from hereditary cases is lacking. A study by Cher *et al.* (8) did indicate that a large portion of the q arm including the 1q24-25 region is frequently increased in copy number in advanced prostate cancer specimens examined by comparative genomic hybridization. Candidate genes in the interval implicated include the *ski*, *abl2*, and *trk* oncogenes as well as *LAMC2*, which encodes an isoform subunit of a basement membrane protein (laminin) (*18*).

The data presented here indicate that a susceptibility locus that may account for a significant fraction of hereditary prostate cancer can be detected in families by linkage analysis. If this linkage is confirmed in an independent data set, then we propose the designation *HPC1* (hereditary prostate cancer 1) for this locus. This observation if confirmed sets the stage for the challenging task of cloning *HPC1* and identifying the responsible genetic alterations in high-risk families. Given that previous segregation analyses have suggested that approximately one in 170 individuals in the United States may carry a dominant susceptibility allele for prostate cancer (6), one can estimate (very roughly) that one in 500 may have an alteration in *HPC1*. Because early diagnosis can be lifesaving in prostate cancer, the potential ability to identify individuals at genetically high risk, especially when combined with methods that detect early signs of malignancy (physical exam, transrectal ultrasound and prostate-specific antigen), could ultimately be of significant medical benefit.

Table 2. Linkage results for susceptibility to prostate cancer and nine markers on chromosome 1 in 91 families. $\hat{Z}$ and $\hat{\Theta}$ represent the maximum lod scores and recombination fractions, respectively. NPL *Z* scores are not directly comparable to parametric $\hat{Z}$ (LOD) scores. Therefore, significance levels are given for the NPL *Z* scores. For parameter (LOD) scores, a *Z* score of 3.0 corresponds to a signifiance level of a 0.0001.

Marker	Distance (cM)*	Parametric analysis: two-point lod		Nonparametric multipoint analysis	
		$\hat{Z}$	$\hat{\Theta}$	*Z* score	*P*
D1S452	—	0.94	0.27	2.28	0.01
D1S218†	1.9	2.31	0.23	2.14	0.02
D1S212	3.6	2.98	0.19	4.22	0.00001
D1S2883	0.0	3.65	0.18	4.16	0.00002
D1S466	5.1	2.41	0.20	4.71	0.000001
D1S2818	0.9	1.69	0.24	4.66	0.000002
D1S158	1.5	2.53	0.21	4.62	0.000002
D1S422	4.4	2.67	0.20	4.26	0.00001
D1S413†	4.9	1.80	0.21	2.83	0.002

*Distances in centimorgans from the preceding marker in the table were derived from the CRIMAP analysis. †Markers used in genome-wide scan.

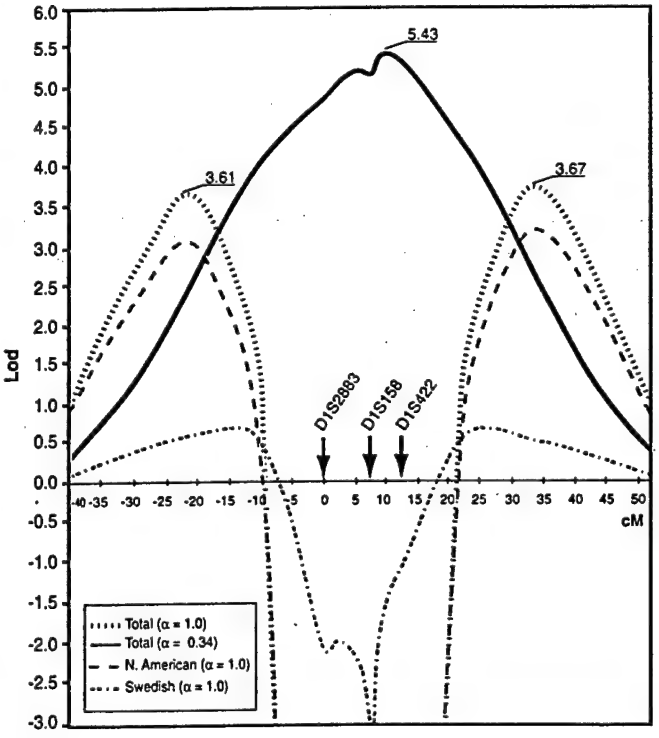


Fig. 2. Multipoint lod scores for the prostate cancer susceptibility locus relative to markers in the 1q24-25 region. Parametric multipoint lod scores were calculated with markers *D1S2883*, *D1S518*, and *D1S422*. The results are plotted as a function of distance from *D1S2883*, and are given for the North American and Swedish families, calculated both independently and combined. The combined values (total) are plotted for values of $\alpha = 1.0$ (assuming all families linked) and for $\alpha = 0.34$ (assuming heterogeneity, with 34% of the families linked). The maximum lod score under homogeneity is 3.67, but it rises to 5.43 if heterogeneity is assumed.

REFERENCES AND NOTES

1. S. L. Parker, T. Tong, S. Bolden, P. A. Wingo, *CA Cancer J. Clin.* **46**, 5 (1996).
2. O. W. Brawley and B. S. Kramer, in *Comprehensive Textbook of Genitourinary Oncology*, N. J. Vogelsang, P. T. Scardino, W. U. Shipley, D. S. Coffey, Eds. (Williams and Wilkins, Baltimore, 1996), p. 565.
3. L. Cannon *et al.*, *Cancer Surveys* **1**, 47 (1982).
4. G. D. Steinberg, B. S. Carter, T. H. Beaty, B. Childs, P. C. Walsh, *Prostate* **17**, 337 (1990).
5. B. S. Carter *et al.*, *J. Urol.* **150**, 797 (1993).
6. B. S. Carter, T. H. Beaty, G. D. Steinberg, B. Childs, P. C. Walsh, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 3367 (1992).
7. T. Visakorpi *et al.*, *Cancer Res.* **55**, 342 (1995); C. Van Den Berg *et al.*, *Clinical Cancer Res.* **1**, 11 (1995).
8. M. L. Cher *et al.*, *Cancer Res.* **56**, 3091 (1996).
9. W. B. Isaacs, in *Genetics and Cancer: A Second Look*, B. A. Ponder, W. K. Cavenee, E. Solomon, Eds. (Cold Spring Harbor Laboratory Press, Plainview, NY, 1995), p. 357.
10. L. A. Cannon-Albright *et al.*, *Am. J. Hum. Genet.* **55**, A147 (1994).
11. D. Freije *et al.*, *Proc. Am. Assoc. Cancer Res.* **37**, 670 (1996).
12. In response to an article in *Parade* magazine (3 March 1996) describing this study, individuals in 1904 different families reported having three or more family members affected with prostate cancer. Of these, 6% reported having five affected family members, 1.4% reported having six affected members, and 1.4% reported having seven or more affected members.
13. North American prostate cancer families were obtained from three sources: 65% of the families were identified by referrals generated as a response to a letter sent by one of us (PCW) to 8000 urologists throughout the country; the second source, accounting for 23% of the families, was identified by family history records of the patient population seen at Johns Hopkins Hospital for treatment of prostate cancer; the remainder of the families responded to articles published in a variety of lay publications describing this study. Prostate cancer diagnosis was verified by medical records for each affected male studied. Swedish families were obtained as a result of a nationwide search of cancer registries, and from referrals from urologists. All individuals in this study gave full informed consent.
14. Genomic DNA was prepared from lymphoblastoid transformed cell lines for the original 66 families, and prepared from whole blood and archived tissue specimens for the additional cohort of 25 families. Overall, samples from 604 individuals were genotyped (339 affected and 265 unaffected individuals); 70 additional unrelated individuals (20 North American and 50 Swedish) were also typed to provide allele frequency estimates for these populations (see *16*). High-throughput, semi-automated genotyping was accomplished by means of ABI 373 and 377 DNA sequencers to resolve multiple, fluorescently labeled markers in each gel lane. An internal size standard enabled allele sizing with the local Southern algorithm in GENESCAN (Applied Biosystems, Foster City, CA). A control individual was typed on each gel as a sizing and binning check. Genotype editing and binning were performed in GENOTYPER (Applied Biosystems, Foster City, CA). All genotyping was done blinded to affected status. A total of 26% of the markers applied were proprietary to the ABI PRISM mapping set; the balance were derived from the Genome Database (Johns Hopkins University School of Medicine, Baltimore, MD). A list of survey markers used will be supplied by the authors upon request. Reverse primer sequences for most markers were modified to promote complete nontem-

plated nucleotide addition to the 3′ end of amplified products by Taq DNA polymerase, enabling reliable identification of 1-base pair alleles present in 7.4% of the markers [M. J. Brownstein, J. D. Carpten, J. R. Smith, *Biotechniques* 20, 1004 (1996)]. We obtained 97.1% of data sought with survey markers. Blinded duplicate typing of 7560 alleles provided a genotyping error rate estimate of 0.26%. The observed rate of non-Mendelian inheritance was 7.06×10^{-4}.

15. In the model used, affected men were assumed to be carriers of a rare autosomal dominant gene frequency $q = 0.003$) (6), with a fixed 15% phenocopy rate, while all unaffected men under 75 and all women were assumed to be of unknown phenotype. In men over age 75, the lifetime penetrance of gene-carriers was estimated to be 63% (based on a population based segregation analysis performed by H.G., in preparation, and the lifetime risk of prostate cancer for non-carriers was 16% in this age class (based on SEER data) [C. L. Kosary, L. A. G. Ries, B. A. Miller, B. F. Hankey, A. Harras, B. K. Edwards (Eds.), SEER Cancer Statistics Review, 1973–1992: Tables and Graphs, National Cancer Institute. NIH Pub. No. 96-2789. Bethesda, MD, 1995]. This is a conservative model as it minimizes the chances of incorrectly assuming that a young unaffected male is a noncarrier. The fact that nonparametric methods produce results of similar statistical significance (Table 2) adds confidence to the conclusion that the observed linkage is not strongly dependent on the choice of this particular model.
16. Standard parametric likelihood analysis was performed by means of FASTLINK [R. W. Cottingham Jr., R. M. Idury, A. A. Schaffer, *Am. J. Hum. Genet.* **53**, 252 (1993)] for two-point linkage and VITESSE [J. R. O'Connel and D. E. Weeks, *Nature Genet.* **11**, 402 (1995)] for multipoint linkage analysis. Multipoint analysis has the advantage of utilizing data from multiple linked markers to maximize the information in a given pedigree. Nonparametric multipoint analysis, which is robust even when the mode of inheritance is not known, was also performed, with GENEHUNTER [L. Kruglayk and E. S. Lander, *Am. J. Hum. Genet.* **57**, 439 (1995)] to calculate normalized Z scores and associated P values. In all of the linkage analyses, allele frequencies for the markers were estimated from independent individuals in the families and unrelated individuals separately for the North American and Swedish families. CRIMAP [E. S. Lander and P. Green, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 2363 (1987)] was used to order the multiple markers on chromosome 1 using the genotype data from all pedigrees. The BUILD option of CRIMAP was first used to establish the order of markers with at least a likelihood ratio of 1000:1. The FLIP option was then used to calculate the likelihood of alternative marker orders by permuting adjacent loci (five flanking markers). The most likely order thus determined is the same as the published order (http://cedar.soton.ac.uk/pub). The admixture test as implemented in HOMOG [J. Ott, *Analysis of Human Genetic Linkage* (Johns Hopkins Univ. Press, Baltimore, 1985), pp. 200–203] was used to test for genetic heterogeneity in the context of the two-point parametric analysis.
17. The evaluation of age as a variable is confounded because of the changing methods used to diagnose this disease, and increased interest in screening for this disease. For the years prior to the use of prostate-specific antigen (PSA), diagnosis of prostate cancer was often not made until men presented with advanced disease, whereas today most men are diagnosed younger and at an earlier stage.
18. The expert technical assistance of C. Ewing and J. Robinson, and the help of X. Chen, D. Schwengel, R. Paul, C. Engstrand, A. Kallioniemi, L. Hardie, and B. Carter during the early phases of this work is acknowledged. We also thank B. Childs, J. Isaacs, and D. Coffey for helpful advice. We acknowledge the assistance of L. Middelton, C. Francomano, and the Family Studies Core of the National Center for Human Genome Research (NCHGR), and the Genetic Resources Core Facility (JHU). We also acknowledge A. Lowe and D. Gilbert at the Applied Biosystems Division of Perkin-Elmer for providing valuable genotyping technical support. We wish to thank all the physicians who referred families for this study. Supported by grants from U.S. Public Health Service SPORE CA58236; The Fund for Research and Progress in Urology; The Johns Hopkins University;. Swedish Cancer Society (Cancerfonden); Lion's Cancer Foundation, Department of Oncology, Umeå Universitet, and a 1995 CaPCURE award. D.F. is supported by a grant from American Foundation for Urologic Disease.

1 October 1996; accepted 24 October 1996

Evidence for a prostate cancer susceptibility locus on the X chromosome

Jianfeng Xu[1*], Deborah Meyers[1], Diha Freije[2], Sarah Isaacs[2], Kathy Wiley[2], Deborah Nusskern[2], Charles Ewing[2], Eric Wilkens[2], Piroska Bujnovszky[2], G. Steven Bova[2,3], Patrick Walsh[2] & William Isaacs[2,4] (*Group 1*)
Johanna Schleutker[5*], Mika Matikainen[5*], Teuvo Tammela[5], Tapio Visakorpi[5] & Olli-P. Kallioniemi[12] (*Group 2*)
Rebecca Berry[6*], Daniel Schaid[7*], Amy French[6], Shannon McDonnell[7], Jennifer Schroeder[6], Michael Blute[8] & Stephen Thibodeau[6] (*Group 3*)
Henrik Grönberg[9*], Monika Emanuelsson[9], Jan-Erik Damber[10], Anders Bergh[11] & Björn-Anders Jonsson[11] (*Group 4*)
Jeffrey Smith[12*], Joan Bailey-Wilson[12], John Carpten[12], Dietrich Stephan[12], Elizabeth Gillanders[12], Isaac Amundson[12], Tommi Kainu[12], Diana Freas-Lutz[12], Agnes Baffoe-Bonnie[13], Anne Van Aucken[12], Raman Sood[12], Francis Collins[12], Michael Brownstein[12] & Jeffrey Trent[12] (*Group 5*)
**These authors contributed equally to this work.*

Over 200,000 new prostate cancer cases are diagnosed in the United States each year, accounting for more than 35% of all cancer cases affecting men, and resulting in 40,000 deaths annually[1]. Attempts to characterize genes predisposing to prostate cancer have been hampered by a high phenocopy rate, the late age of onset of the disease and, in the absence of distinguishing clinical features, the inability to stratify patients into subgroups relative to suspected genetic locus heterogeneity. We previously performed a genome-wide search for hereditary prostate cancer (HPC) genes, finding evidence of a prostate cancer susceptibility locus on chromosome 1 (termed *HPC1*; ref. 2). Here we present evidence for the location of a second prostate cancer susceptibility gene, which by heterogeneity estimates accounts for approximately 16% of HPC cases. This HPC locus resides on the X chromosome (Xq27–28), a finding consistent with results of previous population-based studies suggesting an X-linked mode of HPC inheritance. Linkage to Xq27–28 was observed in a combined study population of 360 prostate cancer families collected at four independent sites in North America, Finland and Sweden. A maximum two-point lod score of 4.60 was observed at *DXS1113*, θ=0.26, in the combined data set. Parametric multipoint and non-parametric analyses provided results consistent with the two-point analysis. Significant evidence for genetic locus heterogeneity was observed, with similar estimates of the proportion of linked families in each separate family collection. Genetic mapping of the locus represents an important initial step in the identification of an X-linked gene implicated in the aetiology of HPC.

Despite the medical significance of prostate cancer in terms of morbidity, mortality and health-care costs, our understanding of the molecular determinants of prostate cancer susceptibility remains rudimentary. Epidemiological studies supporting the existence of hereditary forms of prostate cancer have led to the initiation of genome-wide searches for loci contributing to hereditary prostate cancer. A previous scan for linkage resulted in suggestive evidence (lod>1.0) for prostate cancer susceptibility loci on several chromosomes, including 1q, 4q, 5p, 7p, 13q and Xq (ref. 2). Statistically significant evidence was achieved only for the locus 1q24–25 (*HPC1*). Subsequent stratification of pedigrees showed that families linked to *HPC1* tended to have an early mean age of diagnosis (under 65 years) and a large number of affected members (>4). Even in this subset, this locus accounts for only approximately one-half of the families[3]. Further, although two confirmatory studies have corroborated linkage to *HPC1* (refs 4,5), three additional studies found no clear evidence for *HPC1*-predisposed disease in their study populations[6–8]. The disparity in these studies emphasizes the common set of obstacles for linkage detection in hereditary prostate cancer, most prominently, a high phenocopy rate and genetic locus heterogeneity.

Table 1 • Characteristics of prostate cancer families

	JHU	Mayo	Tampere	Umeå	All
Number of families	139	123	57	41	360
Number of individuals typed	766	407	548	268	1989
Number of affected individuals typed	452	314	137	117	1020
Avg. number of affected/family (range)	5.1 (3–17)	4.0 (3–11)	3.2 (2–9)	4.5 (3–10)	4.3 (2–17)
Avg. number of affected individuals typed/family (range)	3.2 (2–11)	2.6 (2–6)	2.4 (2–9)	2.8 (2–8)	2.7 (2–11)
Avg. age at diagnosis (range)	64.1 (39–85)	67.1 (41–93)	68.2 (45–90)	68.0 (46–86)	66.3 (39–93)

[1]Center for the Genetics of Asthma and Complex Diseases, University of Maryland, Baltimore, Maryland 21201, USA. Departments of [2]Urology, [3]Pathology and [4]Oncology, Johns Hopkins Medical Institutions, Baltimore, Maryland 21287, USA. [5]Laboratory of Cancer Genetics, Institute of Medical Technology, University of Tampere and Tampere University Hospital, Tampere, Finland. Departments of [6]Laboratory Medicine and Pathology, [7]Health Sciences Research and [8]Urology, Mayo Clinic/Foundation, Rochester, Minnesota 55902, USA. Departments of [9]Oncology, [10]Urology & Andrology and [11]Pathology, Umeå University, Umeå, Sweden. [12]Prostate Cancer Investigation Group, National Human Genome Research Institute, National Institute of Health, Bethesda, Maryland 20892, USA. [13]Population Science Division, Fox Chase Cancer Center, Philadelphia, Pennsylvania 19012, USA. Correspondence should be addressed to W.I. (e-mail: wisaacs@jhmi.edu).

Table 2 • Two-point parametric lod scores

Marker	Heterozygosity	cM[b]	lod (θ)[a] JHU (139)	Mayo (123)[c]	Tampere (57)	Umeä (41)[d]	All (360)
DXS984	0.74	140.0	0.40 (0.36)	0.31 (0.34)	0.87 (0.22)	0.03 (0.44)	1.00 (0.34)
DXS1232	0.66	140.9	0.28 (0.36)	0.00 (0.50)	0.66 (0.22)		0.24 (0.40)
DXS1205	0.66	142.3	0.19 (0.38)	0.00 (0.50)	2.05 (0.14)		0.33 (0.36)
DXS6751	0.74	143.6	0.49 (0.36)	0.52 (0.32)	1.56 (0.18)		1.63 (0.32)
DXS6798	0.83	144.8	0.51 (0.36)		0.78 (0.22)		0.87 (0.32)
DXS8106	0.70	146.1	0.82 (0.34)	0.80 (0.30)	0.89 (0.16)		1.93 (0.30)
DXS6806	0.81	147.3	0.45 (0.36)	0.78 (0.30)	0.14 (0.28)	0.03 (0.44)	1.07 (0.34)
DXS8043	0.83	148.8	0.97 (0.32)	0.02 (0.40)	0.00 (0.50)	0.08 (0.38)	0.74 (0.36)
AFMA113zf5	0.68	149.3	0.11 (0.36)	1.24 (0.28)	1.22 (0.18)		2.01 (0.28)
DXS1200	0.60	150.4	1.98 (0.28)	0.86 (0.26)	0.17 (0.32)	0.00 (0.50)	2.80 (0.30)
DXS297	0.70	151.0	0.64 (0.34)	0.18 (0.36)	0.13 (0.00)		0.84 (0.34)
AFM136yb10	0.68	152.5	1.00 (0.30)	0.40 (0.30)	0.05 (0.38)		1.38 (0.32)
DXS8091	0.80	152.5	1.52 (0.30)	0.28 (0.34)	0.00 (0.50)		1.43 (0.32)
DXS1113	0.80	153.0	1.73 (0.28)	1.89 (0.26)	0.49 (0.22)	0.60 (0.26)	4.60 (0.26)
DXS1193	0.78	153.3	0.96 (0.32)		0.58 (0.26)	0.34 (0.32)	1.80 (0.30)
DXS8069	0.67	154.5	0.44 (0.36)	0.84 (0.30)	0.01 (0.40)	0.12 (0.38)	1.20 (0.34)
DXS8011	0.87	154.6	0.32 (0.36)		0.58 (0.26)		0.72 (0.34)
DXS8103	0.77	155.2	0.10 (0.42)	0.38 (0.34)	0.92 (0.24)	0.29 (0.32)	1.10 (0.36)
AFMA225xh9	0.74	156.3	0.31 (0.36)	0.98 (0.30)	0.00 (0.50)		0.68 (0.36)
AFMA08xa5	0.51	157.1	0.02 (0.44)	0.02 (0.40)	0.09 (0.00)		0.03 (0.42)
DXS1108	0.70	158.8	0.12 (0.42)	0.57 (0.32)	0.00 (0.50)		0.42 (0.38)

[a]Maximum lod score under homogeneity with the maximum likelihood estimate of the recombination fraction (θ), calculated using FASTLINK. [b]Distance in cM from Xpter. [c]Three markers were not genotyped in this group. [d]Thirteen markers were not genotyped in this group.

A further confounding issue in prostate cancer linkage studies is the lack of a clear delineation of the mode(s) of inheritance. Segregation analyses of familial prostate cancer have supported an autosomal dominant mode of inheritance for prostate cancer susceptibility alleles[9–11], although formal testing of possible X chromosome segregation has not been performed. On the basis of studies of prostate cancer risk in relatives of affected men, it has been suggested that an HPC susceptibility locus may reside on the X chromosome. Several population-based studies have reported a statistically significant excess risk of prostate cancer in men with affected brothers, as compared with those with affected fathers, consistent with the hypothesis of an X-linked, or recessive, model of inheritance[12–16]. In our initial genome-wide search for prostate cancer linkage, there was suggestive evidence of linkage to the X chromosome[2]. These indications have prompted a more detailed analysis of potential X-linkage in HPC families.

To carry out this analysis, we have assembled 360 prostate cancer pedigrees consisting of families collected at sites in the US (Johns Hopkins University (JHU) in Baltimore, Maryland and the Mayo Clinic in Rochester, Minnesota), Finland (University of Tampere, Tampere) and Sweden (Umeå University, Umeå). Characteristics of the various family collections are given (Table 1). Overall, these 360 families contained 1,568 affected members. DNA samples, either from blood or archival tissue samples, were available from 1,020 affected individuals, and from an additional 969 individuals who were either female or unaffected. Over one-half of the families had at least one case of apparent male-to-male disease transmission. As it is possible that some of these occurrences result from a high phenocopy rate, the entire data set was analysed for possible evidence of X linkage.

The results from our previous 10-cM genome-wide screen using 66 North American prostate cancer families implicated a 40-cM interval from *DXS1001* to *DXS1108*, reaching a maximum two-point lod score of 1.08 at marker *DXS1193* at Xq27–28 (ref. 2). To more rigorously test the hypothesis of linkage to this region, an additional 28 markers were selected to augment the five original survey markers across the X chromosome interval. These markers were genotyped to create density map intervals of 1.2-cM for the 139 North American HPC families collected at JHU. A subset of 26 of these markers, spanning 19 cM from *DXS984* to *DXS1108* (140–159 cM from Xpter), were genotyped for the 123 Mayo Clinic and the 57 Finnish HPC families, and a less dense, 4-cM map of eight markers in this interval was completed for the 41 Swedish families. Allele frequencies were estimated from independent individuals in the complete data set. Two-point parametric lod scores are listed (Table 2). Twelve of the markers tested had lod scores greater than 1 in the combined data set, with a maximum score of 4.6 at marker *DXS1113*, θ=0.26. These results were supported by non-parametric affected sibpair analysis (Table 3). Fourteen consecutive markers had an excess mean identical-by-descent (IBD) sharing (0.55), with the lowest *P*-value of 0.00006 at *DXS1113*. The lod score, on the basis of sibpair IBD sharing,

Table 3 • Two-point affected sibpair analysis

	cM[a]	Mean IBD[b]	*P*-value[c]	lod
DXS984	140.0	0.54	0.08	0.42
DXS1232	140.9	0.51	0.33	0.04
DXS1205	142.3	0.53	0.15	0.24
DXS6751	143.6	0.56	0.005	1.41
DXS6798	144.8	0.55	0.047	0.60
DXS8106	146.1	0.57	0.005	1.43
DXS6806	147.3	0.55	0.039	0.67
DXS8043	148.8	0.55	0.023	0.86
AFMA113zf5	149.3	0.58	0.013	1.08
DXS1200	150.4	0.60	0.00008	3.11
DXS297	151.0	0.56	0.025	0.83
AFM136yb10	152.5	0.57	0.007	1.28
DXS8091	152.5	0.57	0.003	1.63
DXS1113	153.0	0.60	0.00006	3.20
DXS1193	153.3	0.57	0.006	1.37
DXS8069	154.5	0.55	0.048	0.60
DXS8011	154.6	0.55	0.04	0.65
DXS8103	155.2	0.52	0.16	0.20
AFMA225xh9	156.3	0.54	0.06	0.50
AFMA08xa5	157.1	0.52	0.32	0.05
DXS1108	158.8	0.52	0.21	0.14

[a]Distance in cM from Xpter. [b]Affected sibpair analyses were performed using ANALYZE. [c]All possible sibpairs were used in the analysis, however, a weight of (n–1) was given to the sibship of multiple sibs, where n is the number of sibs.

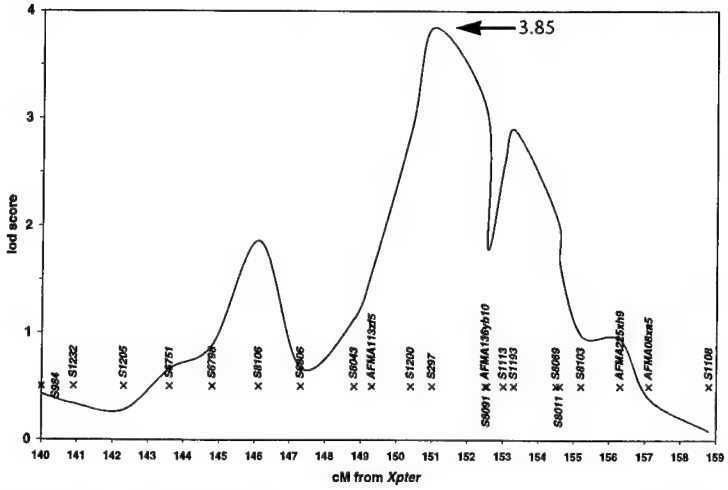


Fig. 1 Graph of multipoint lod scores assuming heterogenity. The peak multipoint lod score of 3.85 is located between *DXS1200* and *DXS297*.

was 3.2 for this marker. When population-specific allele frequencies were used, similar results were obtained.

Simulation studies were performed to estimate the probability of obtaining a two-point parametric lod score of 4.6 or greater, or a *P*-value less than 0.00006 for non-parametric affected sibpair analysis (mean test), at a single marker on the X chromosome in the absence of linkage (false positive rate). Among 10,000 replicates in the simulation, there were no two-point parametric lod scores greater than 4.0, nor were there any *P*-values less than 0.00006 for affected sibpair analysis. There were three lod scores greater than 3, and only once was there a *P*-value less than 0.0001 among the 10,000 replicates.

Results from parametric multipoint linkage analyses were consistent with the two-point analyses. Data from the Swedish families were not included in the multipoint analysis, because only eight markers were genotyped in this dataset. Analysis was carried out using a sliding multipoint approach[17–19], and heterogeneity analysis was then performed using HOMOG (ref. 20). The maximum lod score assuming heterogeneity was 3.85, occurring 151 cM from Xpter, between loci *DXS1200* and *DXS297* (Fig. 1). Significant evidence for locus heterogeneity was obtained, with the proportion (α) of families linked estimated at 16% (χ^2=17.73, df=1, *P*=0.00002; Table 4).

Each study population had positive two-point and multipoint lod scores for multiple markers in the Xq27–28 region (Tables 2,4). Estimates of the proportions of linked families in each collection ranged from 15% (JHU) to 41% (Tampere), although the differences among groups are not statistically significant (χ^2=0.53, *P*=0.77).

As a possible source of genetic heterogeneity, we stratified families into two subsets on the basis of consistency with an X-linked mode of inheritance, using the apparent presence or absence of male-to-male transmission as a single, surrogate, stratification criterion. Following this stratification, 129 families without male-to-male transmission contribute disproportionately to the evidence of linkage to this region (maximum multipoint lod score assuming heterogeneity=2.46 at 151 cM from Xpter, estimated proportion linked=19%). In contrast, for families with male-to-male transmission (n=190), the maximum lod score assuming heterogeneity was 1.47, also at 151 cM, with a lower estimated proportion linked (13%). Although this difference is not statistically significant, the observed trend is consistent with the hypothesis of X chromosome linkage in this data set. The observation of positive lod scores in families with apparent male-to-male disease transmission may result from the presence of phenocopies as affected fathers or other relatives.

As there was evidence for linkage of HPC susceptibility loci to both 1q24–25 (ref. 2) and Xq27–28 in families collected at JHU, we tested the hypothesis (H_1) that there are three types of prostate cancer families in this cohort: (i) a proportion of the families linked to Xq27–28; (ii) a proportion of the remaining families linked to 1q24–25; and (iii) the rest linked to neither region. Using the admixture test[20] (HOMOG3R) with multipoint lod score data for the 139 families in this group, significant evidence of locus heterogeneity was observed (Table 5). The data were made at least 360-fold more probable given the hypothesis (H_1) that subsets of HPC families are linked to Xq27–28 or to *HPC1*, and the remainder unlinked, than the hypotheses of either as a sole locus (H_2 or H_3). Multipoint data suggested that 15% of the families in this group were linked to the X chromosome locus, and that 30% were linked to *HPC1*. Similarly, in the 59 families in this collection that are not linked to *HPC1* (lod<–0.1), the multipoint lod score under heterogeneity is 1.96 for Xq27–28, whereas the lod score is 0.48 in the remaining 80 families.

Linkage analysis is valuable for identification of genetic loci predisposing to prostate cancer. The presence of genetic heterogeneity both in and across populations necessitates large-scale studies to provide significant statistical power to identify major loci. Among the JHU study population, loci at 1q24–25 and Xq27–28 are estimated to account for approximately 30% and 15% of the prostate cancer families, respectively. In contrast, of these two loci, only the X-chromosome locus appears to have a prominent role in prostate cancer predisposition in the Finnish study population, in which a larger fraction of families (over 40%) are estimated to be X-linked, and *HPC1* shows only a marginal role (J. Schleutker *et al.*, in preparation). A similar situation exists in the Mayo Clinic data set, although the proportion of families linked to the X chromosome is the same as in the JHU study population. From these results, we anticipate that confirmatory studies will also encounter genetic heterogeneity. Indeed, a recently described factor contributing to the lack of linkage to *HPC1* in several family collections may be the presence of an increased proportion of X-linked pedigrees in these cohorts. Similarly, linkage to the X chromosome may be most readily apparent upon stratification of pedigrees by male-to-male disease transmission in these populations, although, as we have seen, evidence for this linkage is not restricted to particular subsets of this stratification. Further, as the major proportion of the families examined in this study are not linked to either *HPC1* or the X-chromosome locus, and as collection of additional study

Table 4 • Heterogeneity test using multipoint lod score for each family collection

		MLE estimates (3-unit support interval)		
Group	lod[a]	α	map position[b]	*P*-value
JHU	2.34	0.15 (0.03–0.30)	152.5 (140.0–154.6)	0.001
Mayo	1.03	0.16 (0.01–0.34)	154.5 (140.0–158.8)	0.029
Tampere	2.03	0.41 (0.08–0.71)	143.6 (140.0–151.0)	0.002
All	3.85	0.16 (0.06–0.26)	151.0 (140.0–153.3)	0.00002

[a]Heterogeneity test was based on sliding multipoint lod scores, using the admixture test (HOMOG), where lod is calculated assuming heterogeneity. [b]Distance in cM from Xpter.

Table 5 • Admixture test using multipoint lod scores for Xq27–28 and 1q24–25 (139 JHU families)

Hypothesis[a]	% linked to 1q24–25[b] α1	% linked to Xq27–28[c] α2	ln L	χ^2 (df)[d]	*P*-value
H1	0.30	0.15	16.43		
H2	0.29	[0]	10.52	11.82 (2)	0.0027
H3	[0]	0.16	5.42	22.02 (2)	1.6×10^{-5}
H4	[0]	[0]	0	32.86 (5)	4.0×10^{-7}

[a]Heterogeneity test was performed using the admixture test (HOMOG3R). [b]Multipoint lod scores at 1q24–25 were based on markers *D1S158* and *D1S422*. [c]Multipoint lod score at Xq27–28 were based on markers AFMA113zf5, *DXS1200* and *DXS297*. [d]χ^2 is –2ln likelihood difference between H1 and each alternative hypotheses.

populations increases the statistical power, additional loci may be proven to account for a portion of prostate cancer predisposition. In this regard, a recent study of 47 French and German families had a multipoint lod score, assuming heterogeneity, of 2.2 (α=50%) and two-point score of 2.7 at 1q42.2–43 (ref. 8).

Significance of the results achieved here is supported by several different lines of evidence. Most importantly, the linkage data derived from each of four independent family collections provides evidence of linkage to Xq27–28. When combined, this data set yields a maximum two-point lod score of 4.6, meeting the proposed criteria for significant linkage[21]. Second, non-parametric methods supported this result and provided a model-independent significance level of *P*=0.00006 for linkage. Third, simulations performed to provide an empirical nominal significance level for the observed linkage results never yielded a two-point lod score greater than 4.0, nor any *P*-value less than 0.00006 in 10,000 replicates. The data support the newly identified locus as predisposing to hereditary prostate cancer at Xq27–28.

A candidate prostate cancer susceptibility gene residing on the X chromosome is the androgen receptor gene[22–25] (*AR*). *AR*, however, is located at Xq12, over 50 cM from the region implicated in this study. Furthermore, direct assessment of linkage to *AR* in the North American families studied here provides no evidence of linkage (unpublished observations). Several genes at Xq27–28 have been mapped (http://www.ncbi.nlm.nih.gov/genemap), and these and other novel genes in the Xq27–28 region will need to be evaluated as candidate prostate cancer susceptibility genes.

We have presented evidence for linkage of a significant subset of prostate cancer families to a locus on Xq27–28. Contingent upon confirmation, we suggest the designation *HPCX* for this locus.

Methods

North American families. Johns Hopkins family collection: The 79 North American families that were described in the report of linkage to *HPC1* (ref. 2) are included in this study, as are an additional 60 pedigrees collected at the Brady Urologic Institute at Johns Hopkins. A majority of these families were ascertained through referrals from physicians; some families were recruited from earlier epidemiological studies[9] and through news articles. Age of diagnosis of prostate cancer was confirmed either through medical records or from two other independent sources. All individuals in this study gave full informed consent.

Mayo Clinic family collection: The 123 North American families in this collection were ascertained by a cancer family-history survey, sent to over 5,000 men who underwent a radical prostatectomy for clinically localized prostate cancer in the Department of Urology at the Mayo Clinic during 1966–1995 (ref. 11). Prostate cancer diagnosis and the age of onset was confirmed through medical records at the Mayo Clinic and elsewhere. All participants in this study gave full informed consent.

Finnish families. In Finland, 302 prostate cancer families with two or more affected cases were identified through referrals from physicians, family questionnaires sent to patients, a nationwide registry-based search and advertisements in newspapers, radio and television. Of this group, 57 families that were informative for linkage analyses were included in this study. Diagnosis of all prostate cancer patients was confirmed through hospital records or from the Finnish cancer registry. All individuals participating in this study gave full informed consent.

Swedish families. Since 1995, families with three or more relatives affected with prostate cancer have been collected at the Department of Oncology of Umeå University, mainly from referrals from urologists throughout Sweden. From approximately 300 referrals, 41 families informative for linkage analysis have been selected. Twelve of these families were included in an earlier report[2]. When blood samples were unavailable, tissue samples were collected from affected men whenever possible. Tissue samples were reviewed by an experienced pathologist and microdissection was performed to separate normal and tumour tissue. For genotyping, only normal tissue was used. All prostate cancer diagnoses in the families were confirmed by the National cancer registry and medical records.

Genotyping methods. Techniques of preparing DNA and genotyping were as described[2]. Markers were derived from the Genome Database (Johns Hopkins University School of Medicine). Marker data was obtained for 33 polymorphic loci available in the GDB, spanning the approximately 40-cM interval between *DXS1001* and *DXS1108*. Order and distance for these markers was estimated from the entire genotype data set using CRIMAP (ref. 26). The most likely order thus determined agrees with the published order[27]. Allele frequencies were estimated from genotypes of independent individuals in the 360 families.

Statistical methods. Both parametric and non-parametric linkage approaches were used in this study. The parametric analysis used a previous model[2,3] with regard to disease allele frequency (0.003) and age-specific penetrances, although affected men were assumed to be carriers of an X-linked, sex-limited, dominant gene. A fixed 15% phenocopy rate, that is, P (non-predisposing genotype/disease), was assumed, whereas all unaffected men under 75, and all women, were assumed to be of unknown phenotype. In men over age 75, the lifetime penetrance of gene carriers was estimated to be 63%, and the lifetime risk of prostate cancer for a non-carrier was 16% in this age class. FASTLINK (refs 18,19) and ANALYZE (ftp://linkage.cpmc.columbia.edu/software/analyze) were used for the parametric two-point analysis. For the non-parametric analysis, affected sibpairs were used for the two-point analysis as implemented by ANALYZE, using the mean test and likelihood based test. The mean test compares the number of alleles shared IBD with the number of alleles not shared IBD among affected sibpairs. When there are multiple sibs in a sibship, a weight of (n–1) is given to the sibship, where n is the number of sibs. When parents are not genotyped, the program computes the likelihood of each possible genotype for the parents, and computes the number of alleles shared IBD in a sibpair as the average over all possible parental genotype combinations, weighted by their conditional probabilities given the known data.

The simulation study was performed using FASTSLINK (ftp://watson.hgen.pitt.edu/pub). A 10-allele marker, which represents the marker *DXS1113*, was simulated unlinked to the disease locus using the exact pedigree structure and availability of genotype information for the 360 families analysed. The marker *DXS1113* has 15 alleles, six of which have frequencies of approximately 1% or less. To make the simulation of a large number of replicates (10,000) more practical, we collapsed the six less frequent alleles into one allele.

The multipoint approach is critical in linkage analysis of a late age-of-onset disease such as prostate cancer, because parental genotypic data are often missing, making inference of IBD ambiguous. Additionally, multipoint analysis is more robust to misspecification of allele frequencies and statistical fluctuations at individual loci. When more markers are used simultaneously in the analysis (multipoint analysis), the probability distribution is concentrated on certain inheritance vectors, thus the determination of IBD is less dependent on the marker allele frequencies[28]. However, multipoint analyses of X-chromosome marker data are hampered by the lack of fully functional X-chromosome versions of the most appropriate multipoint analysis computer programs (for example, GENEHUNTER).

In this study, the parametric multipoint analysis was performed using FASTLINK (LINKMAP; refs 18,19). Due to computer memory constraint, only 4-point analyses (disease locus against three marker loci) were performed. A sliding multipoint approach was used as described[17]. Briefly, this approach consists of sliding a group of three loci down the map and analysing the disease locus only in the interval between the second and third marker. Heterogeneity analysis was then performed using HOMOG (ref. 20).

The admixture model was used to test several hypotheses for genetic locus heterogeneity (HOMOG3R; ref. 20). α_1 is the proportion of families linked to the first disease locus (that is, 1q24–25), and α_2 is the proportion linked to the second disease locus (that is, Xq27–28). Hypothesis 1 (H_1) assumes that there are three types of families in the sample, (α_1, α_2 and $1-(\alpha_1+\alpha_2)$). Hypothesis 2 (H_2) assumes that there are two types of families, α_1 and $1-\alpha_1$. Hypothesis 3 (H_3) assumes that there are two types of families, α_2 and $1-\alpha_2$. Hypothesis 4 (H_4) assumes no linkage to either disease locus ($\alpha_1=\alpha_2=0$). Maximum likelihood for each of these hypotheses was calculated from the data. Chi-square (χ^2) tests were performed by calculating twice the difference of the natural log likelihood between two hypotheses, with the degrees of freedom (df) equal to the difference in the number of parameters estimated for the two hypotheses. The asymptotic null distribution of the test statistic has not been well investigated, but this approach is conservative[20].

Stratification of families. The criteria used to categorize a family as having evidence of male-to-male disease transmission were as follows: (i) presence of affected father and affected son(s) combinations, or (ii) prostate cancer case(s) on the paternal side of the family, with no evidence of affected relatives on maternal side. Families that did not meet these criteria were classified as families without evidence of male-to-male transmission.

Acknowledgements

The authors thank all the family members who participated in this study. In addition, the assistance of physicians who referred patients to this study is gratefully acknowledged. The authors wish to thank K. Vaherto, M. Kuorilehto, R. Sankila, E. Pukkala and J. Viitanen for assistance in family collection in Finland. This work was supported by the National Institutes of Health (SPORE CA58236, CA72818 and NO1-HG-55389); The Fund for Research and Progress in Urology, The Johns Hopkins University; Academy of Finland; Reino Lahtikari and Sigrid Juselius Foundations; Finnish Cancer Society; Medical Research Fund of Tampere University Hospital; Swedish Cancer Society (Cancerfonden); Lion's Cancer Foundation, Department of Oncology, Umeå University; and CaPCURE. S.I. is supported by the GE Fund. D.F. is supported by the American Foundation for Urologic Disease.

Received 3 June; accepted 1 September, 1998.

1. Landis, S.H., Murray, T., Bolden, S. & Wingo, P.A. Cancer statistics, 1998. *CA Cancer J. Clin.* **48**, 6–29 (1998).
2. Smith, J.R. *et al.* Major susceptibility locus for prostate cancer on chromosome 1 suggested by a genome-wide search. *Science* **274**, 1371–1374 (1996).
3. Grönberg, H. *et al.* Early age at diagnosis in families proving evidence of linkage to the hereditary prostate cancer locus *(HPC1)* on chromosome 1. *Cancer Res.* **57**, 4707–4709 (1997).
4. Cooney, K.A. *et al.* Prostate cancer susceptibility locus on chromosome 1q: a confirmatory study. *J. Natl Cancer Inst.* **89**, 955–959 (1997).
5. Hsieh, C.L. *et al.* Prostate cancer susceptibility locus on chromosome 1q: a confirmatory study. *J. Natl Cancer Inst.* **89**, 1893–1894 (1997).
6. McIndoe, R.A. *et al.* Linkage analysis of 49 high-risk families does not support a common familial prostate cancer-susceptibility gene at 1q24–25. *Am. J. Hum. Genet.* **61**, 347–353 (1997).
7. Eeles, R.A. *et al.* Linkage analysis of chromosome 1q markers in 136 prostate cancer families. *Am. J. Hum. Genet.* **62**, 653–658 (1998).
8. Berthon, P. *et al.* Predisposing gene for early-onset prostate cancer, localized on chromosome 1q42.2–43. *Am. J. Hum. Genet.* **62**,1416–1424 (1998).
9. Carter, B.S., Beaty, T.H., Steinberg, G.D., Childs, B. & Walsh P.C. Mendelian inheritance of familial prostate cancer. *Proc. Natl Acad. Sci. USA* **89**, 3367–3371 (1992).
10. Grönberg, H., Damber, L., Damber, J.E. & Iselius, L. Segregation analysis of prostate cancer in Sweden: support for dominant inheritance. *Am. J. Epidemiol.* **146**, 552–557 (1997).
11. Schaid, D.J., McDonnell, S.K., Blute, M.L. & Thibodeau, S.N. Evidence for autosomal dominant inheritance of prostate cancer. *Am. J. Hum Genet.* **62**, 1425–1438 (1998).
12. Woolf, C.M. An investigation of the familial aspects of carcinoma of the prostate. *Cancer* **13**, 361–364 (1960).
13. Narod, S.A. *et al.* The impact of family history on early detection of prostate cancer. *Nature Med.* **1**, 99–101 (1995).
14. Whittemore, A.S. *et al.* Family history and prostate cancer risk in black, white, and Asian men in the United States and Canada. *Am. J. Epidemiol.* **141**, 732–740 (1997).
15. Monroe, K.R. *et al.* Evidence of an X-linked or recessive genetic component to prostate cancer risk. *Nature Med.* **1**, 827–829 (1995).
16. Hayes, R.B. *et al.* Prostate cancer risk in U.S. blacks and whites with a family history of cancer. *Int. J. Cancer* **60**, 361–364 (1995).
17. Terwilliger, J.D. & Ott, J. *Handbook of Human Genetic Linkage.* (Johns Hopkins University Press, Baltimore, 1994).
18. Cottingham Jr, R.W., Idury, R.M. & Schaffer, A.A. Faster sequential genetic linkage computations. *Am. J. Hum. Genet.* **53**, 252–263 (1993).
19. Lathrop, G.M., Lalouel, J.M., Julier, C. & Ott, J. Strategies for multilocus linkage analysis in humans. *Proc. Natl Acad. Sci. USA* **81**, 3443–3446 (1984).
20. Ott, J. *Analysis of Human Genetic Linkage* (Johns Hopkins University Press, Baltimore, 1991).
21. Lander, E.S. & Kruglayk, L. Genetic dissection of complex traits: guideline for interpreting and reporting linkage results. *Nature Genet.* **11**, 241–247 (1995).
22. Stanford, J.L. *et al.* Polymorphic repeats in the androgen receptor gene: molecular markers of prostate cancer risk. *Cancer Res.* **57**, 1194–1198 (1997).
23. Hakimi, J.M., Rondinelli, R.H., Schoenberg, M.P. & Barrack, E.R. Androgen-receptor gene structure and function in prostate cancer. *World J. Urol.* **14**, 329–337 (1996).
24. Giovannucci, E. *et al.* The CAG repeat within the androgen receptor gene and its relationship to prostate cancer. *Proc. Natl Acad. Sci. USA* **94**, 3320–3325 (1997).
25. Irvine, R.A., Yu, M.C., Ross, R.K. & Coetzee, G.A. The CAG and GGC microsatellites of the androgen receptor gene are in linkage disequilibrium in men with prostate cancer. *Cancer Res.* **55**, 1937–1940 (1995).
26. Lander, E.S. & Green, P. Construction of multilocus genetic linkage maps in humans. *Proc. Natl Acad. Sci. USA* **84**, 2363–2367 (1987).
27. Collins, A., Teague, J., Keats, B.J. & Morton, N.E. Linkage map integration. *Genomics* **36**, 157–162 (1996).
28. Kruglayk, L., Daly, M.J., Reeve-Daly, M.P. & Lander, E.S. Parametric and nonparametric linkage analysis: A unified multipoint approach. *Am. J. Hum. Genet.* **58**, 1347–1363 (1996).

Hum Genet (2001) 108:335–345
DOI 10.1007/s004390100488

ORIGINAL INVESTIGATION

Jianfeng Xu · Siqun L. Zheng · Bao-li Chang
Jeffrey R. Smith · John D. Carpten · O. Colin Stine
Sarah D. Isaacs · Kathy E. Wiley · Lauren Henning
Charles Ewing · Piroska Bujnovszky
Eugene R. Bleeker · Patrick C. Walsh · Jeffrey M. Trent
Deborah A. Meyers · William B. Isaacs

Linkage of prostate cancer susceptibility loci to chromosome 1

Received: 2 January 2001 / Accepted: 12 February 2001 / Published online: 28 March 2001

Abstract Three prostate cancer susceptibility genes have been reported to be linked to different regions on chromosome 1: HPC1 at 1q24–25, PCAP at 1q42–43, and CAPB at 1p36. Replication studies analyzing each of these regions have yielded inconsistent results. To evaluate linkage across this chromosome systematically, we performed multipoint linkage analyses with 50 microsatellite markers spanning chromosome 1 in 159 hereditary prostate cancer families (HPC), including 79 families analyzed in the original report describing HPC1 linkage. The highest lod scores for the complete dataset of 159 families were observed at 1q24–25 at which the parametric lod score assuming heterogeneity (hlod) was 2.54 (P=0.0006) with an allele sharing lod of 2.34 (P=0.001) at marker D1S413, although only weak evidence was observed in the 80 families not previously analyzed for this region (hlod=0.44, P=0.14, and allele sharing lod=0.67, P=0.08). In the complete data set, the evidence for linkage across this region was very broad, with allele sharing lod scores greater than 0.5 extending approximately 100 cM from 1p13 to 1q32, possibly indicating the presence of multiple susceptibility genes. Elsewhere on chromosome 1, some evidence of linkage was observed at 1q42–43, with a peak allele sharing lod of 0.56 (P=0.11) and hlod of 0.24 (P=0.25) at D1S235. For analysis of the CAPB locus at 1p36, we focused on six HPC families in our collection with a history of primary brain cancer; four of these families had positive linkage results at 1p36, with a peak allele sharing lod of 0.61 (P=0.09) and hlod of 0.39 (P=0.16) at D1S407 in all six families. These results are consistent with the heterogeneous nature of hereditary prostate cancer, and the existence of multiple loci on chromosome 1 for this disease.

J. Xu · E. R. Bleeker · D. A. Meyers
Center for Human Genomics,
Wake Forest University School of Medicine, Winston-Salem, N.C., USA

S. L. Zheng · B.-L. Chang · O. C. Stine
Department of Epidemiology,
University of Maryland School of Medicine, Baltimore, MD, USA

J. R. Smith · J. D. Carpten · J. M. Trent
National Human Genome Research Institute,
National Institute of Health, Bethesda, MD, USA

S. D. Isaacs (✉) · K. E. Wiley · L. Henning · C. Ewing
P. Bujnovszky · P. C. Walsh · W. B. Isaacs
Department of Urology, Johns Hopkins Medical Institutions,
Baltimore, MD, USA
e-mail: wisaacs@jhmi.edu,
Tel.: +1-410-9552518, Fax: +1-410-9550833

W. B. Isaacs
Marburg 115, Johns Hopkins Hospital, 600 N. Wolfe Street,
Baltimore, MD 21287, USA

Present address:
J. R. Smith
Vanderbilt University Medical Center,
Division of Genetic Medicine, MRB II, 442B, 2200 Pierce Avenue,
Nashville, TN 37232–6304, USA

Introduction

Three prostate cancer (MIM 176807) susceptibility loci have been reported to be linked to three different regions on chromosome 1 (Smith et al. 1996; Berthon et al. 1998; Gibbs et al. 1999). By studying 79 hereditary prostate cancer (HPC) families (defined as three or more prostate cancer patients in first-degree relatives) ascertained at Johns Hopkins Hospital and 12 HPC families ascertained in Sweden, Smith et al. (1996) reported the first prostate cancer linkage to markers at 1q24–25 (HPC1; MIM 601518). The peak two-point lod score was 3.65 at a recombination fraction (θ) of 0.18 with marker D1S2883. Multipoint analyses with various combinations of three consecutive markers were performed, and lod scores greater than 4 were observed. Significant evidence for locus heterogeneity was obtained by an admixture test with the proportion of linked families (α) estimated to be 34%. The maximum multipoint lod score under the assumption of heterogeneity was 5.43. Non-parametric analyses provided comparable results, with a peak multipoint NPL score of 4.71 (P=1E-5). The linkage was stronger in the subset of families with early age of onset (Grönberg et al. 1997) and in families with

evidence of male-to-male disease transmission (Xu et al. 1998; Xu and ICPCG 2000).

The results of analysis of HPC1 linkage by other research groups have been variable. Several independent studies corroborated linkage to HPC1. Cooney et al. (1997) reported a linkage study of 1q24–25 in 59 prostate cancer families, each with two or more affected individuals. The peak NPL score was 1.58 at D1S466 (P=0.057) in the total 59 families but was 1.72 (P=0.045) in the subset of 20 families that met the criteria for hereditary prostate cancer families (three or more affected individuals within one nuclear family, affected individuals in three successive generations, and/or clustering of two or more individuals affected <55 years). Hsieh et al. (1997) provided further evidence in support of HPC1. In 92 unrelated families having three or more affected individuals, the NPL score was 1.71 (P=0.046). The evidence for linkage was stronger in the 46 families with a mean age at diagnosis of less than 67 years. The NPL score was 2.04 (P=0.023). Neuhausen et al. (1999) presented positive evidence for linkage in 41 large HPC families ascertained in Utah. The peak two-point lod was 1.73 (P=0.005) in the total families and a two-point lod of 2.82 (P=0.0003) in early age of onset families. Finally, in a study of 144 HPC families collected at Mayo Clinic, Berry et al. (2000a) did not find evidence for linkage at HPC1 region in the total sample but established HPC1 linkage in a subset of 102 families with male-to-male disease transmission. The peak NPL score was 1.99 (P=0.03) at D1S212.

Four other groups, however, reported no significant evidence for linkage of HPC1 in their study populations. McIndoe et al. (1997) found no evidence for linkage in this region in 49 high-risk prostate cancer families, with either a parametric lod score approach assuming homogeneity or a non-parametric analysis. There was also no evidence for linkage in the 18 families with early age at diagnosis (<65 years). Linkage analysis was further extended to 150 HPC families in this study population, and the linkage to HPC1 was strongly rejected (Goode et al. 2000). Berthon et al. (1998) reported results of a genome-wide screen and specific results from the 1q24–25 region in 47 French and German families. For the three markers in the 1q24–25 region, they found negative two-point lod scores assuming a dominant model. Eeles et al. (1998) published a linkage study of 1q24–25 in 136 prostate cancer families ascertained in United Kingdom, Quebec, and Texas, 76 of which had three or more affected individuals. They found negative NPL scores in this region in the total sample but positive NPL scores in a subset of 35 families with four or more affected members. Suarez et al. (2000a) obtained no evidence for the HPC1 locus in their 230 multiplex sibships, although positive linkage results in the region were observed. The Zlr was 2.10 (P=0.018) at D1S2141 in sibships with positive family history and Zlr=2.72 (P=0.003) at D1S1677 in sibships with negative family history. Suarez et al. (2000b) reported further negative findings for HPC1 in their 45 new multiplex sibships and four expanded families.

To clarify the inconsistent replication results and to test for linkage in a larger data set, a combined analysis for six markers in the 1q24–25 regions was performed in 772 HPC families ascertained by members of the International Consortium for Prostate Cancer Genetics (ICPCG) from North America, Australia, Finland, Norway, Sweden, and the United Kingdom (Xu and ICPCG 2000). This group of families included the majority of families analyzed in the studies described above but did not include the original 91 families described by Smith et al. (1996) in which the original linkage to HPC1 was found. Overall, there was some evidence for linkage, with a peak parametric multipoint lod score assuming heterogeneity (hlod) of 1.40 (P=0.01) at D1S212. The estimated α was 6%. The evidence for linkage was stronger in families with male-to-male disease transmission. The peak hlod was 2.56 (P=0.0006), and an α of 11% was seen in the subset of 491 families with male-to-male disease transmission families, compared with hlod of 0 in the remaining 281 families. Within the male-to-male disease transmission families, the α increased with early mean age of diagnosis (<65, α=19%) and number of affected family members (≥5, α=15%). The highest α was observed for the 48 families that met all three criteria (peak hlod =2.25, P=0.001, α=29%). The results from non-parametric analyses were consistent with the parametric analysis, with a peak NPL score of 1.14 at D1S212 in the total 772 HPC families. The strongest evidence for linkage at this region was observed in the 491 families with male-to-male disease transmission, with a peak NPL of 2.3 (P=0.01).

These results support the finding of a prostate cancer susceptibility gene linked to 1q24–25.

The second HPC locus (PCaP; MIM 602759) on chromosome 1 was reported in the data from 47 French and German HPC families by using the combination of genome-wide screening and fine mapping (Berthon et al. 1998). This locus is located at 1q42–43, which is about 60 cM from HPC1. The maximum two-point lod score was 2.7 at marker D1S2785. The multipoint parametric analysis yielded an hlod of 2.2, and the non-parametric multipoint analysis yielded an NPL score of 3.1 (P=0.001). The estimated proportion of linked families was 50% in the sample. Furthermore, the evidence for linkage was stronger in a subset of nine families with early mean age of onset (<60 years), with hlod of 3.31 and NPL of 3.32 (P=0.001). However, replication studies of this locus have yielded inconsistent results in other study populations. Gibbs et al. (1999) analyzed 152 HPC families by using markers spanning a 20-cM region of 1q42–43 and did not find evidence for linkage to an HPC susceptibility locus. No evidence for linkage was found in early age of onset families. The most suggestive evidence for linkage was found in subset of 38 families with five and more affected members, with NPL of 1.2 (P=0.1). Whittemore et al. (1999) evaluated linkage to the region by using three markers in 97 HPC families. Negative lod scores and NPL scores were observed in the total sample and in 48 early age of onset families and 49 late age of onset families. Berry et al. (2000a) did not find evidence for linkage at six markers at 1q42–43 region in either the total 144 HPC families studied or in the subset of early age of onset families. However, they

found suggestive evidence for linkage in 21 families that met all three criteria: male-to-male disease transmission, family mean age of onset <66, and more than five affected members. The peak NPL score was 1.45 (P=0.08).

The third HPC susceptibility locus on chromosome 1 (PCBP/CAPB; MIM 603688; at 1p36) was reported by Gibbs et al. (1999) in families with prostate cancer and brain cancer. Based on the data from an initial genome-wide screen in 70 HPC families, evidence for linkage was observed at 1p36, with a multipoint hlod of 1.65 and NPL score of 2.13 (P=0.02). A fine mapping study was then performed in the region with additional markers and an additional 71 HPC families. Stronger evidence for linkage in the region was seen in a subset of 12 families with a history of prostate cancer and primary brain cancer. The overall two-point lod score was 3.22 at D1S507 in this subset. In the younger age of onset group of six HPC families (mean age at diagnosis <66 years), a maximum two-point lod of 3.65 at D1S407 was observed. The peak multipoint lod score assuming heterogeneity was 0.81 in the six families. No evidence for linkage was seen in either early or late age of onset families without a history of brain cancer. To replicate the finding in an independent study population, Berry et al. (2000a) studied 13 HPC families with prostate cancer and brain cancer and found no evidence for linkage. Both multipoint lod scores and NPL scores were negative in the region. Badzioch et al (2000) found evidence of linkage to CAPB in families with early onset prostate cancer, although no association with other cancers was seen.

Other prostate cancer linkages located outside of chromosome 1 have been reported. In a linkage analysis of combined data of 360 prostate cancer families from North America, Finland, and Sweden, Xu et al. (1998) reported evidence for a prostate cancer susceptibility locus on Xq27–28 (HPCX; MIM 300147), with a maximum two-point lod of 4.6 at DXS1113. Parametric and non-parametric multipoint analyses provided results consistent with the two-point analysis. Stratified analysis on the basis of consistency with an X-linked mode of inheritance revealed that 129 families without male-to-male disease transmission contributed disproportionately to the evidence of linkage to this region. The other prostate cancer susceptibility locus resided at chromosome 20q13 (HPC20). It was identified in 162 North American families with three or more members affected with prostate cancer (Berry et al. 2000b). The highest two-point lod score was 2.69 at D20S196, and the maximum multipoint NPL score was 3.02 (P=0.002) at D20S887. The evidence for linkage at this region was stronger in subsets of families with male-to-male disease transmission, with fewer than five family members affected with prostate cancer, and with later average age of diagnosis (≥66 years). Recently, several genome-wide scans in prostate cancer families have been reported that implicate a number of novel loci as harboring prostate cancer susceptibility loci (Gibbs et al. 2000; Suarez et al. 2000a; Witte et al. 2000).

In light of the three reported prostate cancer susceptibility loci on chromosome 1 and the inconsistent results from replication studies, we systematically evaluated the linkage of prostate cancer susceptibility loci to the three proposed regions on chromosome 1 by using a dense marker set spanning the entire chromosome. We studied 159 HPC families ascertained at Johns Hopkins Hospital, including 79 described previously by Smith et al. (1996). This study had the following three specific goals: (1) to test for linkage(s) of prostate cancer susceptibility loci across the complete length of chromosome 1, especially with regard to (a) the linkage at 1q24–25 in the subset of 80 new families and linkage in the complete family collection, (b) the linkage at 1q42–43 in the complete family collection, and (c) the linkage at 1p36 in 6 families with history of both prostate cancer and primary brain cancer; (2) to investigate the relationship of the three reported linkages on chromosome 1 to one another; and (3) to perform stratified analyses to explore characteristics of the families supporting these linkages in terms of male-to-male disease transmission, mean age of onset within a family, and number of affected members.

Methods

Family collection

All 159 HPC families were collected and studied at the Brady Urology Institute at Johns Hopkins Hospital (Baltimore, Md.). The first 79 HPC families had been included in the initial HPC1 report (Smith et al. 1996), and the remaining 80 families were recruited subsequently. Families were ascertained from three resources. Most of them were ascertained through referrals generated as a response to a letter by one of us (P.C.W.) to 8000 urologists throughout the country. The second source was identified from family history records of the patient population seen at Johns Hopkins Hospital for treatment of prostate cancer. The remaining families came from respondents to articles published in a variety of lay publications describing our prostate cancer family studies. Prostate cancer diagnosis was verified by medical records for each affected male studied. Age of diagnosis of prostate cancer was confirmed either through medical records or from two other independent sources. All individuals in this study gave full informed consent.

Families were defined as having male-to-male disease transmission when there was evidence of paternal disease transmission in the families, including the following: (1) affected father and affected sons; (2) prostate cancer cases on the paternal side of the family, with no evidence of affected relatives on the maternal side; or (3) prostate cancer cases on the maternal side of the family and male-to-male disease transmission on the maternal side. The remaining families were defined as non-male-to-male disease transmission families. They had either an unknown mode of inheritance (insufficient data to determine inheritance pattern) or were consistent with an X-linked mode of inheritance.

The family characteristics of the 159 HPC families are shown in Table 1. The subsequently collected 80 HPC families tended to be smaller and more heterogeneous in terms of race/ethnicities, compared with the first 79 HPC families. The classification of the number of affected family members was based on their medical history, and not all affected members had DNA samples. Fourteen and eleven of the families in the complete data set were African-American and Ashkenazi Jewish, respectively.

Genotyping and markers

Fifty microsatellite markers across chromosome 1 were genotyped and analyzed for the study. These markers were selected based on

Table 1 Characteristics of prostate cancer families

	All families	1st 79 families[a]	2nd 80 families
Mean age at onset (years)	64.3	65.1	63.5
Mean number of affected family members	5.1	5.3	4.9
Mean number of affected family members with DNA sample	3.3	3.8	2.9
Male-to-male disease transmission			
No. families with male-to-male disease transmission	99 (62%)	47 (59%)	52 (65%)
No. families without male-to-male disease transmission	60	32	38
Age of onset			
No. families age onset <65	79 (50%)	35 (44%)	44 (55%)
No. families age onset >= 65	80	44	36
No. families with >=5 affected members	90 (57%)	48 (61%)	42 (53%)
No. families with 4 affected members	40	23	17
No. families with 3 affected members	29	8	21
Race/ethnicity			
Caucasian	133 (84%)	74 (94%)	59 (75%)
African American	14	3	11
Others	12	2	10

[a]These families were included in the initial linkage report of HPC1 locus (Smith et al. 1996)

the following three criteria: (1) in the regions where linkages were reported (1p36, 1q24–25, and 1q42–43), polymorphic markers were selected with a resolution of approximately 2 cM; (2) the markers with the highest lod scores in each of the three initial reports were selected; and (3) in the regions in between these three reported linkages, markers were selected with a resolution about 10 cM. We performed multiplex polymerase chain reaction (PCR) with fluorescently labeled primers (either fam, hex, or ned), and the resulting PCR fragments were separated by using capillary electrophoresis in a ABI 3700 sequencer. The genotypes were scored by using ABI software (Genotyper). A modified version of the program Linkage Designer (http://dnalab-www.uia.ac.be/dnalab/ld.html) binned the alleles and checked inheritance. The output from Linkage Designer was then analyzed further for any inconsistencies by running LINKAGE software (Lathrop et al. 1984; Cottingham et al. 1993) without disease information. The Hardy-Weinberg equilibrium test was performed by using computer program GDA (Weir et al. 1996) for all markers as another check for the quality of genotype. The final check that was performed on the data was to run CRIMAP (Lander and Green 1987) to determine the order and length of the chromosomal map and to detect double recombinants. Marker allele frequencies were estimated from the independent individuals in the data set. Marker order and distances were based on the Marshfield genetic map (Table 2). In the cases where the markers were unavailable from a Marshfield map, the order and distances were estimated from the data by using CRIMAP. We chose the markers D1S489 and D1S552 as the boundaries of 1p36 region because these markers and markers in between had NPL scores of more than 1 in the original report (Gibbs et al. 1999). Markers D1S452 and D1S249 were chosen as the boundaries of 1q24–25 region because these markers and markers in between had NPL scores of more than 2 in the original report (Smith et al. 1996). For the boundaries of the 1q42–43 region, we chose markers D1S251 and D1S2842 as these markers flank the region of positive two-point lod scores in the original report (Berthon et al. 1998).

Statistical analyses

Multipoint linkage analyses were performed by using both parametric and non-parametric methods, implemented by the computer program GENEHUNTER-PLUS (Kruglyak et al. 1996; Kong and Cox 1997). For the parametric analysis, the same autosomal dominant model that had been used in many of the previous prostate linkage studies was assumed (Smith et al. 1996; Berthon et al. 1998). Under this model, the disease gene frequency of 0.003, incomplete penetrance, and phenocopies were assumed. Specifically, affected men were assumed to be disease gene carriers, with a fixed 15% phenocopy rate, whereas all unaffected men under 75 and all women were assumed to be of unknown phenotype. In men aged over 75 years, the lifetime penetrance of gene carriers was estimated to be 63%, and the lifetime risk of prostate cancer for non-carriers was 16% in this age class. Linkage in the presence of heterogeneity was assessed by use of Smith's admixture test for heterogeneity (Ott 1998). In this test, two types of families were assumed, one type linked to the disease locus with a proportion of α, and the other type is not linked with the proportion $1-\alpha$. A maximum likelihood approach was used to estimate the proportion of linked families (α), by maximizing the admixed lod score (hlod).

For the non-parametric analysis, the estimated marker identical by descent (IBD) sharing of alleles for the various affected relative pairs was compared with its expected values under the null hypothesis of no linkage. A statistic "Z-all" in the program was used (Whittemore and Halpern 1994). Allele sharing lod scores were then calculated based on the statistic "Z-all" and assigning equal weight to all families by using the computer program ASM (Kong and Cox 1997).

Both hlod and allele sharing lod can be converted to a χ^2 (χ^2=4.6×hlod). Although the true distribution of the χ^2 under null hypothesis of no linkage is unknown, especially in the situation of multipoint analysis, we assume that the distribution is a mixture of one that is degenerate at zero, and one that can be approximated by the distribution of the maximum of two independent χ^2 variables, each with 1 degree of freedom (Faraway 1993). *P*-values were thus calculated by $0.5*(1-(1-p_1)(1-p_1))$, where p_1 is the *P*-value of χ^2 with 1 degree of freedom.

Linkage analyses conditional on the linkage results at other locations were used in the current study for two purposes. First, for the chromosomal regions that are unlinked but located adjacent to one another (for example, 1q24–25 and 1q42–43), conditional analysis was used to explore whether the evidence for linkage in families linked at one region (the conditional locus) extended to the other region (the test locus). In this case, in the analysis of linkage data for the test locus, a weight of 1 was assigned to families with positive linkage scores at the conditional locus, and families with zero or negative linkage scores at this locus were assigned a weight of 0. Second, for the chromosomal regions that were completely unlinked, conditional analysis was used to explore the interaction of two regions of linkage, either assuming heterogeneity interaction (families linked to one region do not link to another region) or multiplicative interaction (families linked to one region tend to

Table 2 Marker information

Markers	Distance	Heterozygotsity
D1S489	30	0.88
D1S402	31.1	0.94
D1S407	33.9	0.88
D1S3669	37.1	0.91
D1S552	45.4	0.88
D1S1622	55.8	0.92
D1S3728	89.6	0.95
D1S1665	102.1	0.80
D1S1728	109.1	0.86
D1S1588	125.6	0.86
D1S223	133.9	0.73
D1S1631	137	0.91
D1S248	139.1	0.80
D1S2809	144.5	0.76
D1S534	151.5	0.92
D1S514	152	0.65
HSD3B2	152.5	0.74
D1S1653	164.7	0.88
D1S2707	169.1	0.83
D1S1677	176.2	0.89
D1S2799	183.8	0.92
D1S1619	188.9	0.89
D1S452	189.4	0.93
D1S218	192.1	0.94
D1S2659	192.7	0.90
D1S212	194.4	0.94
D1S2883	195.5	0.92
D1S466	198.9	0.93
D1S2818	199	0.92
D1S158	200.6	0.94
D1S191	201.6	0.91
D1S2848	201.7	0.93
D1S202	202.2	0.91
D1S238	203.3	0.94
D1S422	206	0.89
D1S2757	209.8	0.91
D1S413	213.1	0.93
D1S249	221.2	0.93
D1S425	231.7	0.90
D1S2141	234	0.93
D1S399	240.3	0.93
D1S549	240.4	0.89
D1S251	245.6	0.94
D1S235	255.2	0.92
D1S2678	256.9	0.89
D1S2670	263.6	0.92
D1S2785	266.9	0.92
D1S321	268.1	0.90
D1S304	268.2	0.80
D1S2842	274.2	0.90

linked to another region). For the multiplicative interaction, the same weighting scheme mentioned above was used. For the heterogeneity interaction, families were assigned a weight of 1 if they had negative linkage scores at the conditional locus and a weight of 0 if they had zero or positive linkage scores at this locus.

Results

Multipoint linkage analysis with 50 markers spanning chromosome 1

Fifty markers spanning chromosome 1 were genotyped in 159 HPC families, and the data analyzed using both a parametric model and a non-parametric allele-sharing approach. The lod score curves are shown in Fig. 1. The strongest evidence for linkage in the complete data set was observed at D1S413 at 1q25. Evidence for linkage extended from this marker almost 100 cM proximally, as far as 1p13. Additional smaller peaks were detected at D1S3728 at 1p32 and at D1S235 at 1q42.

Analysis of HPC1

The marker D1S413 is located in the region previously defined as HPC1 by Smith et al. (1996). The hlod was 2.54 (P=0.0006), with an estimated α of 17%, and the allele sharing lod was 2.34 (P=0.001) for this marker (Fig. 1). In the 80 new HPC families, the evidence for linkage at this region is substantially weaker (hlod=0.44, P=0.14, and allele sharing lod=0.67, P=0.08) when compared with results from the 79 families described in the original report of HPC1 linkage (hlod=3.05, P=0.0002, and allele sharing lod=3.09, P=0.0002). The evidence for linkage across the region was very broad, with allele sharing lod scores greater than 0.5 extending 100 cM, flanked by markers D1S514 at 1p13 and D1S2141 at 1q32.

Analysis of PCaP

There was evidence for linkage at 1q42–43, but this did not reach statistical significance. The highest allele sharing lod and hlod were 0.56 (P=0.11) and 0.24 (P=0.25) at D1S235, respectively (Fig. 1). This latter marker was at approximately 255 cM from 1pter, located at the proximal boundary of the initially reported PCaP region.

Analysis of CAPB

Although there was no evidence for linkage at 1p36 in the complete set of families (Fig. 1), four of the six families with a history of both prostate cancer and primary brain cancer had positive linkage scores at 1p36 (PCBP/CAPB). The highest allele sharing lod and hlod in the region were 0.61 (P=0.09) and 0.39 (P=0.16) at D1S407, respectively, in the six families. Lod scores at 1q24–25 and 1q42–43 for these six families were zero throughout these regions.

Analysis of 1p32

A linkage signal approximately 85 cM from 1pter was observed in this analysis. The hlod was 0.93 (P=0.04), and the

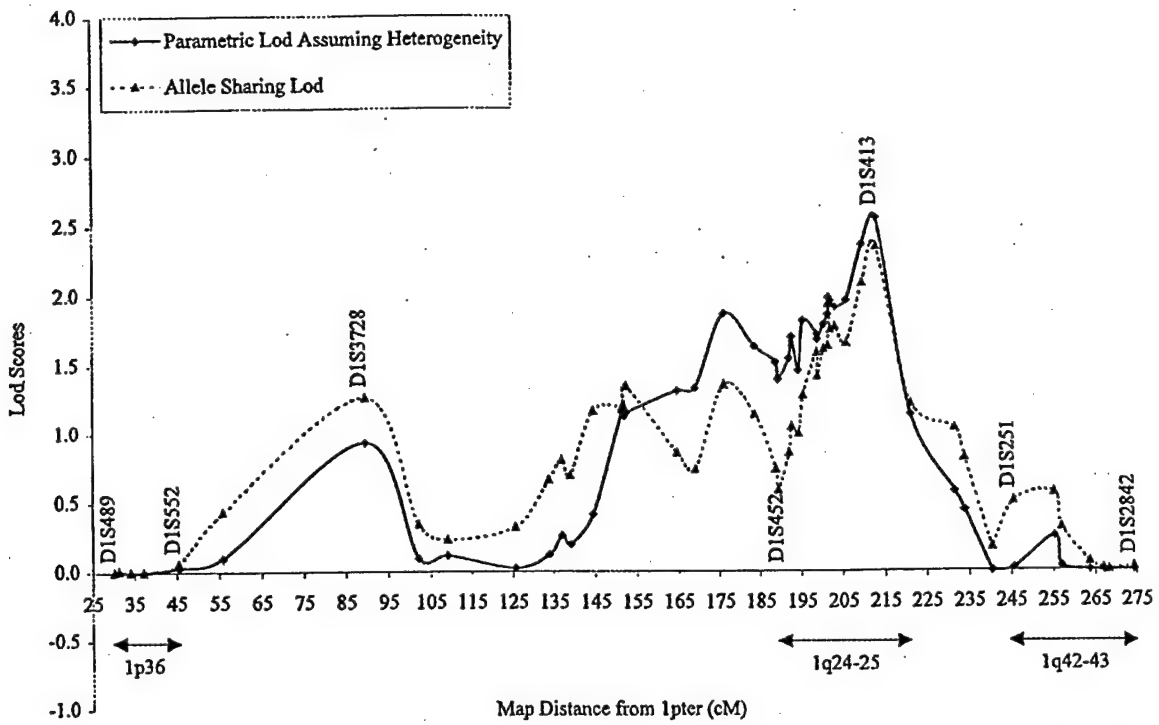


Fig. 1 Results of multipoint parametric and non-parametric linkage analyses of prostate cancer susceptibility loci by using 50 markers across chromosome 1 in 159 hereditary prostate cancer families (*solid line* parametric lod assuming heterogeneity, *dotted line* allele sharing lod, *diamonds* positions of markers, *circles* positions of markers)

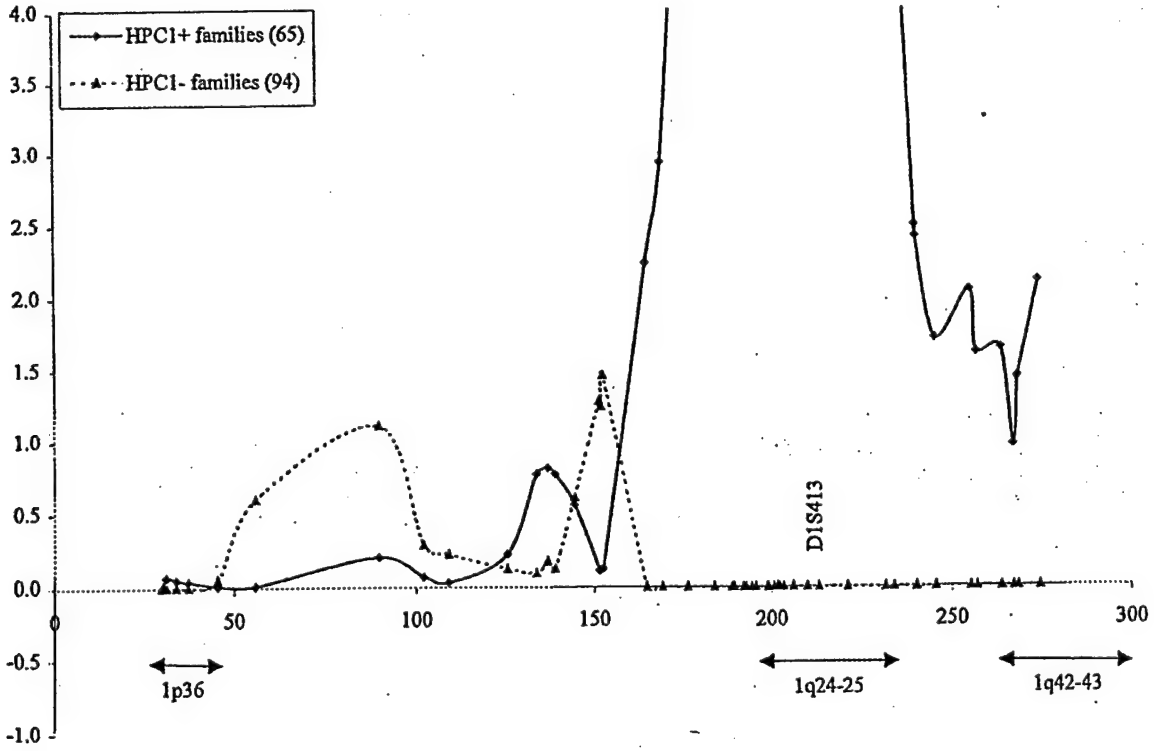


Fig. 2 Results of multipoint allele sharing lod conditional on the linkage result at D1S413 (*solid line* linkage results by assigning a weight of 1 or 0 for families that had allele sharing lod >0 or ≤0 at D1S413, respectively, *dotted line* linkage results by assigning a weight of 1 or 0 for families that had allele sharing lod <0 or ≥0 at D1S413, respectively)

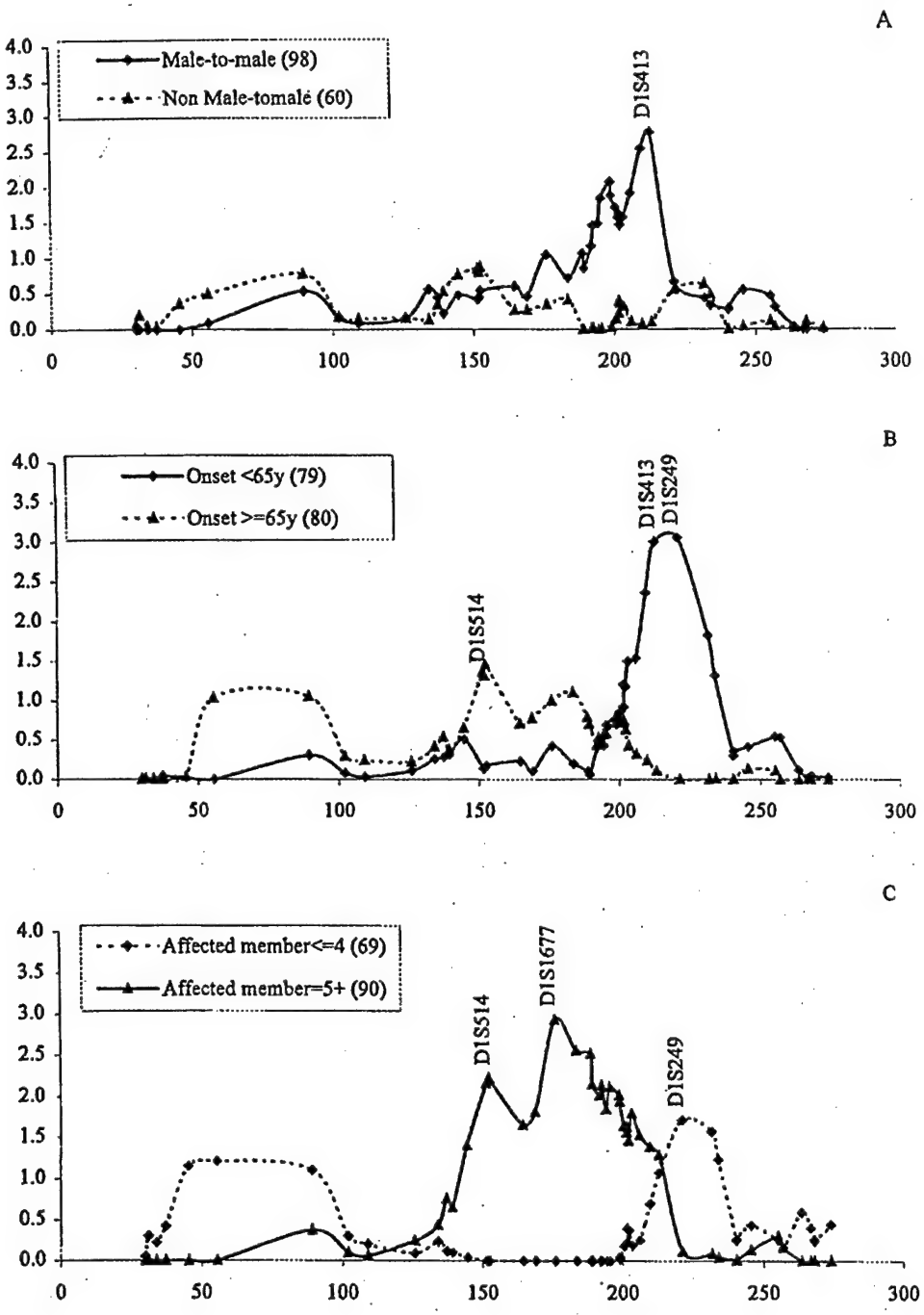


Fig.3A–C Results of multipoint allele sharing lod for 50 markers on chromosome 1 stratified by the family characteristics. **A** Stratification by male-to-male disease transmission. **B** Stratification by family mean age of onset. **C** Stratification by number of affected family members

allele sharing lod was 1.26 (P=0.02) at marker D1S3728 at 1p32. However, these results should be considered preliminary until additional markers in the region are analyzed.

Relationship between linkage at 1q24–25 and other chromosome 1 loci

Since the strongest evidence for linkage was at 1q24–25, linkage data were re-analyzed for chromosome 1 markers conditional on the linkage information at 1q24–25 (Fig. 2). For 1q42–43, the evidence for linkage increased when families having a positive allele sharing lod at marker D1S413 at 1q24–25 region (n=65) were assigned a weight of 1 in the analysis, and the remaining families (n=94) were assigned a weight of 0. The allele sharing lod was 2.26 at D1S235 under these conditions, compared with 0.56 in the unconditional analysis. The results suggested that, in most families linked to 1q24–25, the evidence for linkage extended to markers in the 1q42–43 region. Testing for the independence of the allele sharing lod scores by family be-

tween the regions at D1S413 and D1S235 showed significant dependence between the two regions, with χ^2_1=17.27 (P=0.00003), again indicating that the families linked to 1q24–25 tended to be linked to 1q42–43, and vice versa. It is important to note that the largely inflated lod scores at 1q24–25 are artificial, since families linked to the region were assigned a weight of 1 and families unlinked to the 1q24–25 were assigned a weight of 0; therefore the value of the lod score for the region is not interpretable.

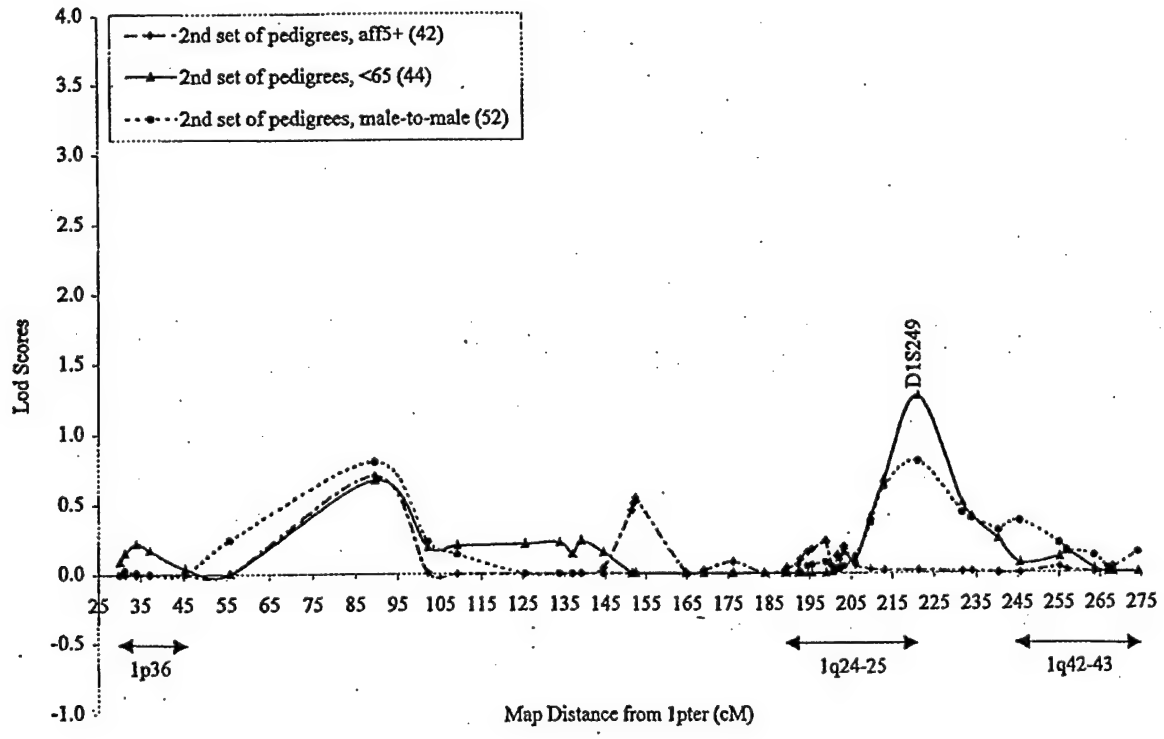
Conversely, when the 65 families that had positive allele sharing lod scores at marker D1S413 were assigned a weight of 0 and the remaining families were assigned a weight of 1, no evidence for linkage at 1q42–43 was observed, and hence little evidence for linkage at 1q42–43 in families not linked to 1q24–25. However, a linkage peak was observed approximately 155 cM from 1pter (1p13) under this conditional analysis (Fig. 2). The peak allele sharing lod was 1.46 (P=0.009) at D1S514. Testing for the independence of the allele sharing lod scores by family between the regions at D1S413 and D1S514 yield a χ^2_1=3.45 (P=0.06). These results suggested that the evidence for linkage at marker D1S514 and D1S413 came from different families.

Stratified analyses of linkage data based on family characteristics

Additional multipoint linkage analyses for all 50 markers across chromosome 1 were performed to include stratification of families based on the presence of male-to-male disease transmission, mean age of onset, and number of affected members (Fig. 3). Both parametric and non-parametric analyses gave similar results; thus, only the results of non-parametric analyses are shown. When the analyses were stratified by the presence or absence of male-to-male disease transmission, evidence for linkage was observed primarily at 1q24–25, occurring in the 98 families with male-to-male disease transmission. The peak lod score was 2.76 (P=0.0004) at D1S413. No statistically significant evidence for linkage was observed in any region in the remaining 60 families without male-to-male disease transmission (Fig. 3A). When families were divided by mean age of onset, the 79 families with early age of onset (<65) provided disproportional evidence for linkage at 1q24–25, with a peak lod of 3.05 (P=0.0002) between D1S413 and D1S249 (Fig. 3B). The 80 families with later age of onset had much weaker evidence for linkage in the region, with the peak lod of 1.45 (P=0.01) at D1S514. When families were stratified by the number of affected members, the 90 families with at least five affected members provided the strongest evidence for linkage in a broad region between 145 cM and 210 cM. The peak lod was 2.93 (P=0.0002) at D1S1677 (Fig. 3C). In the families with fewer than five affected members, evidence for linkage was weaker at D1S249. The peak lod was 1.71 (P=0.005).

The same stratification linkage analyses for the entire region on chromosome 1 were performed for the subset of the new 80 HPC families (Fig. 4). The 44 early age onset families provided the strongest evidence for linkage at the 1q24–25 region, with a peak allele sharing lod of 1.26 (P=0.02) at D1S249. The 52 male-to-male disease trans-

Fig. 4 Results of stratified multipoint allele sharing lod for 50 markers on chromosome 1 in the subset of 80 new HPC families (*aff5+* five or more affected family members)



mission families provided disproportional evidence for linkage with allele sharing lod of 0.8 (P=0.05) at the same marker. Families with at least five affected members did not provide evidence for linkage at 1q24–25.

Discussion

Multipoint linkage analyses for prostate cancer susceptibility loci by using markers across chromosome 1 in 159 HPC families provided several findings of interest. First, the most significant evidence for linkage was observed at 1q24–25 in the complete data set, although the evidence for linkage from the subset of the 80 new families analyzed was weak. The evidence for linkage in this region spanned a broad interval, extending between 1p13 and 1q32. Second, a positive but not statistically significant linkage was observed at 1q42–43. Third, in six families with both prostate cancer and primary brain cancer patients, there was positive linkage at 1p36. Fourth, the evidence for linkage at 1q24–25 mainly came from a subset of families with male-to-male disease transmission and early age of onset.

Since 79 of the 159 families were included in the original report of linkage at 1q24–25 (Smith et al. 1996), the suggestive evidence for linkage at 1q24–25 in the current study cannot be interpreted as an independent confirmation but rather as a further evaluation of linkage in a larger sample. The independent confirmation of the linkage at the region, from the 80 new families, was weak with a peak hlod of 0.44 and an allele sharing lod of 0.67. The reasons for the different levels of support for the linkage in the initial 79 families and in the subsequent 80 families are unknown and could be attributable to a number of factors. (1) Most (70%) of the patients in the second cohort of families were diagnosed in 1992 or later and many of them through prostate-specific antigen (PSA) screening, whereas in the initial 79 families, only 46% were diagnosed by 1992 or later, and fewer were detected by PSA. The year and method of diagnosis could have an impact on the linkage results, probably by affecting the rates of phenocopies (Xu et al. 2000). (2) There are different degrees of genetic locus heterogeneity in the two sets of collected families.

In retrospect, it is possible that the proportion of families linked to 1q24–25 was over-estimated in the initial report (Smith et al. 1996); this is a common phenomenon in initial reports of linkage. Any linkage peak is likely to be at least the combination of two factors: the "true" evidence for linkage to a disease susceptibility gene in some families and the evidence for linkage observed attributable to the random variation by chance in other families (Suarez et al. 1994; Kruglyak et al. 1996). The random variation in favor of linkage may disappear in replication studies or, at the other extreme, result in decreased evidence for linkage. One approach to decrease the impact of random variation and to obtain a reliable estimate is to perform linkage in a large sample. This has been achieved in a combined data analysis of 1q24–25 from the ICPCG group (Xu and ICPCG 2000). The ICPCG study has replicated the linkage in an independent collection of 772 families and provided an estimate that 9% of HPC families are linked to 1q24–25 in the 863 HPC families that were available at the time (including the 79 HPC families and another 12 Swedish HPC families included in the initial finding by Smith et al. 1996).

The chromosomal region with evidence for linkage at 1q24–25 extends across a large genomic interval (~100 cM). The size of this region suggests the presence of multiple prostate cancer susceptibility genes in this interval. Preliminary evidence to support this possibility has been provided by the conditional linkage analyses and χ^2 tests, which indicate independence of the linkages to 1p13 and 1q24–25, i.e., different families are linked to different regions. The presence of multiple genes within this region could partially explain the difficulties experienced in the past 4 years by groups attempting to clone the HPC1 gene.

This is our first report of replication results of the linkage at 1q42–43 in this family collection. Although the linkage results at 1q42–43 are not statistically significant, our results are consistent with a prostate susceptibility locus (PCaP) in the 1q42–43 region. However, further studies with conditional analysis and the χ^2 test for the independence of lod score by families between the regions of 1q42–43 and 1q24–25 suggest that the evidence for both regions is related. A large fraction of families linked to 1q24–25 extend their linkage to 1q42–43. Regarding the previously reported characteristics of families linked to the PCAP locus (Berthon et al. 1997), the evidence for linkage was not increased in the 79 early age of onset families. The highest allele sharing lod was 0.53 (p=0.11) in this group.

This is also our first report of replication results of linkage at 1p36 in our family collection. With only six families with a history of both prostate cancer and primary brain cancer available for testing, we do not have an adequate sample size to make a reliable inference. However, the limited results from the current study are consistent with a prostate susceptibility locus (PCBP/CAPB) in the region. Two of the families had a mean age of onset of less than 65 years, and three of the families had five or more affected family members. In this small group, we have not observed a trend of increased evidence for linkage in the subset of early age of onset in these families, as indicated by the study of Gibbs et al. (1999).

There were a small number of African-American families (n=14) and Ashkenazi Jewish families (n=11) in our study sample. Both groups of families provided some evidence for linkage at 1q24–25, with allele sharing lod scores at D1S413 of 0.53 (P=0.11) and 0.70 (P=0.07), respectively. This compares with a lod score of 2.02 (P=0.002) at this marker for the 133 Caucasian families. Since racial differences in the marker allele frequencies are likely to exist between Caucasian and African-Americans, and linkage analysis is susceptible to the estimates of marker allele frequencies because of the missing parental data, we repeated the analysis for the African-American group by using the marker allele frequencies estimates from individuals in the 14 African-American families. The allele sharing lod was 0.49 at the same marker. No evidence for linkage at 1q42–43 or 1p36 was observed in the African-

American families. For the Ashkenazi Jewish families, allele sharing lod scores of 0.95 (P=0.04) at D1S3669 and 1.31 (P=0.014) at D1S2670 were observed in the CAPB and PCAP regions, respectively. The only Ashkenazi Jewish family with a history of both prostate and primary brain cancer gave a lod score of 0.29 in the CAPB region. Although these results are of interest, the small number of families in both these racial groups stresses the need for cautious interpretation of the data and for larger follow-up studies.

Prostate cancer is a complex disease with many factors that can potentially affect linkage studies. However, considering the public health significance of the disease, the continued evidence for an important role of genetic and familial factors (Carter et al. 1992; Lichtenstein et al. 2000) and the approaches available for mapping disease genes, e.g., the linkage-based positional cloning approach, represent important and potentially productive avenues for investigating and characterizing this common disease.

Acknowledgements The authors thank all the family members who participated in this study. This work was partially supported by PHS SPORE CA58236 and a grant from the Department of Defense to W.B.I.

Electronic-Database Information

- Online Mendelian Inheritance in Man (OMIM): http://www.ncbi.nml.nih.gov/Omim (prostate cancer, MIM 176807; HPC1, MIM 601518; PCaP, MIM 602759; PCBP/CAPB, MIM 603688; HPCX, MIM 300147
- GDA: Software for the Analysis of Discrete Genetic Data: http://lewis.eeb.uconn.edu/lewishome/gda.html
- Linkage Designer: http://dnalab-www.uia.ac.be/dnalab/ld.html

References

Badzioch M, Eeles R, Leblanc G, Foulkes WD, Giles G, Edwards S, Goldgar D, Hopper JL, Bishop DT, Moller P, Heimdal K, Easton D, Simard (2000) Suggestive evidence for a site specific prostate cancer gene on chromosome 1p36. The CRC/BPG UK Familial Prostate Cancer Study Coordinators and Collaborators. The EU Biomed Collaborators. J Med Genet 37:947–949

Berry R, Schaid DJ, Smith JR, French AJ, Schroeder JJ, McDonnell SK, Peterson BJ, Wang ZY, Carpten JD, Roberts SG, Tester DJ, Blute ML, Trent JM, Thibodeau SN (2000a) Linkage analyses at the chromosome 1 loci 1q24–25 (HPC1), 1q42.2–43 (PCAP), and 1p36 (CAPB) in families with hereditary prostate cancer. Am J Hum Genet 66:539–546

Berry R, Schroeder JJ, French AJ, McDonnell SK, Peterson BJ, Cunningham JM, Thibodeau SN, Schaid DJ (2000b) Evidence for a prostate cancer-susceptibility locus on chromosome 20. Am J Hum Genet 67:82–91

Berthon P, Valeri A, Cohen-Akenine A, Drelon E, Paiss T, Wohr G, Latil A, et al (1998) Predisposing gene for early-onset prostate cancer, localized on chromosome 1q42.2–43. Am J Hum Genet 62:1416–1424

Carter BS, Beaty TH, Steinberg GD, Childs B, Walsh PC (1992) Mendalian inheritance of familial prostate cancer. Proc Natl Acad Sci USA 89:3367–3371

Cooney KA, McCarthy JD, Lange E, Huang L, Miesfeldt S, Monite JE, Oesterling JE, et al (1997) Prostate cancer susceptibility locus on chromosome 1q: a confirmatory study. J Natl Cancer Inst 89:955–959

Cottingham RW Jr, Idury RM, Schaffer AA (1993) Faster sequential genetic linkage computations. Am J Hum Genet 53:252–263

Eeles RA, Durocher F, Edwards S, Teare D, Badzioch M, Hamoudi R, Gill S, et al (1998) Linkage analysis of chromosome 1q markers in 136 prostate cancer families. The Cancer Research Campaign/British Prostate Group U.K. Familial Prostate Cancer Study Collaborators. Am J Hum Genet 62:653–658

Faraway JJ (1993) Distribution of the admixture test for the detection of linkage under heterogeneity. Genet Epidemiol 10:75–83

Gibbs M, Stanford JL, McIndoe RA, Jarvik GP, Kolb S, Goode EL, Chakrabarti L, Schuster EF, Buckley VA, Miller EL, Brandzel S, Li S, Hood L, Ostrander EA (1999) Evidence for a rare prostate cancer-susceptibility locus at chromosome 1p36. Am J Hum Genet 64:776–787

Gibbs M, Stanford JL, Jarvik GP, Janer M, Badzioch M, Peters MA, Goode EL, Kolb S, Chakrabarti L, Shook M, Basom R, Ostrander EA, Hood L (2000) A genomic scan of families with prostate cancer identifies multiple regions of interest. Am J Hum Genet 67:100–109

Goode EL, Stanford JL, Chakrabarti L, Gibbs M, Kolb S, McIndoe RA, Buckley VA, Schuster EF, Neal CL, Miller EL, Brandzel S, Hood L, Ostrander EA, Jarvik GP (2000) Linkage analysis of 150 high-risk prostate cancer families at 1q24–25. Genet Epidemiol 18:251–275

Grönberg H, Xu J, Smith JR, Carpten JD, Isaacs SD, Freije D, Bova GS, et al (1997) Early age of diagnosis in families providing evidence of linkage to the hereditary prostate cancer locus (HPC1) on chromosome 1. Cancer Res 57:4707–4709

Hsieh CL, Oakley-Girvan I, Gallagher RP, Wu AH, Kolonel LN, Teh CZ, Halpern J, et al (1997) Re: prostate cancer susceptibility locus on chromosome 1q: a confirmatory study. J Natl Cancer Inst 89:1893–1894

Kong A, Cox NJ (1997) Allele-sharing models: LOD scores and accurate linkage tests. Am J Hum Genet 61:1179–1188

Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES (1996) Parametric and nonparametric linkage analysis: a unified multipoint approach. Am J Hum Genet 58:1347–1363

Lander ES, Green P (1987) Construction of multilocus genetic linkage maps in humans. Proc Natl Acad Sci USA 84:2363–2367

Lathrop GM, Lalouel JM, Julier C, Ott J (1984) Strategies for multilocus linkage analysis in humans. Proc Natl Acad Sci USA 81:3443–3446

Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M, Pukkala E, Skytthe A, Hemminki K (2000) Environmental and heritable factors in the causation of cancer – analyses of cohorts of twins from Sweden, Denmark, and Finland. N Engl J Med 343:78–85

McIndoe RA, Stanford JL, Gibbs M, Jarvik GP, Brandzel S, Neal CL, Li S, et al (1997) Linkage analysis of 49 high-risk families does not support a common familial prostate cancer-susceptibility gene at 1q24–25. Am J Hum Genet 61:347–353

Neuhausen SL, Farnham JM, Kort E, Tavtigian SV, Skolnick MH, Cannon-Albright LA (1999) Prostate cancer susceptibility locus HPC1 in Utah high-risk pedigrees. Hum Mol Genet 8:2437–2442

Ott J (1998) Analysis of human genetic linkage, 3rd edn. Johns Hopkins Press. Baltimore

Smith JR, Freije D, Carpten JD, Grönberg H, Xu J, Issacs SD, Brownstein MJ, Bova GS, et al (1996) Major susceptibility locus for prostate cancer on chromosome 1 suggested by a genome-wide search. Science 274:1371–1374

Suarez BK, Hampe CL, Van Eerdewegh P (1994) Problems of replicating linkage claims in psychiatry. In: Gershon ES, Cloninger CR (eds) Genetic approaches to mental disorders. American Psychiatric Press, Washington DC, pp 23–46

Suarez BK, Lin J, Burmester JK, Broman KW, Weber JL, Banerjee TK, Goddard KA, Witte JS, Elston RC, Catalona WJ (2000a) A genome screen of multiplex sibships with prostate cancer. Am J Hum Genet 66:933–944

Suarez BK, Lin J, Witte JS, Conti DV, Resnick MI, Klein EA, Burmester JK, Vaske DA, Banerjee TK, Catalona WJ (2000b) Replication linkage study for prostate cancer susceptibility genes. Prostate 45:106–114

Weir BS (1996) Genetic data analysis. II. Methods for discrete population genetic data. Sinauer Association, Sunderland, Mass.

Whittemore A, Halpern J (1994) A class of tests for linkage using affected pedigree members. Biometrics 50:118–127

Whittemore AS, Lin IG, Oakley-Girvan I, Gallagher RP, Halpern J, Kolonel LN, Wu AH, Hsieh CL (1999) No evidence of linkage for chromosome 1q42.2–43 in prostate cancer. Am J Hum Genet 65:254–256

Witte JS, Goddard KA, Conti DV, Elston RC, Lin J, Suarez BK, Broman KW, Burmester JK, Weber JL, Catalona WJ (2000) Genomewide scan for prostate cancer-aggressiveness loci. Am J Hum Genet 67:92–99

Xu J, ICPCG (2000) Combined analysis of hereditary prostate cancer linkage to 1q24–25: results from 772 hereditary prostate cancer families from the International Consortium for Prostate Cancer Genetics. Am J Hum Genet 66:945

Xu J, Meyers DA, Freije D, Isaacs S, Wiley K, Nusskern D, Ewing C, et al (1998) Evidence for a prostate cancer susceptibility locus on the X chromosome. Nat Genet 20:175–179

Xu J, Isaacs SD, Wiley KE, et al (2000) Prostate cancer linkage to *HPC1*: effect of year of diagnosis. J Urol 163:55

Am. J. Hum. Genet. 69:341–350, 2001

Linkage and Association Studies of Prostate Cancer Susceptibility: Evidence for Linkage at 8p22-23

Jianfeng Xu,[1] Siqun L. Zheng,[2] Gregory A. Hawkins,[1] Dennis A. Faith,[3] Brian Kelly,[3] Sarah D. Isaacs,[3] Kathleen E. Wiley,[3] Bao-li Chang,[2] Charles M. Ewing,[3] Piroska Bujnovszky,[3] John D. Carpten,[4] Eugene R. Bleecker,[1] Patrick C. Walsh,[3] Jeffrey M. Trent,[4] Deborah A. Meyers,[1] and William B. Isaacs[3]

[1]Center for Human Genomics, Wake Forest University School of Medicine, Winston-Salem, NC; [2]University of Maryland School of Medicine, and [3]Department of Urology, Johns Hopkins Medical Institutions, Baltimore; and [4]National Human Genome Research Institute, National Institutes of Health, Bethesda, MD

Multiple lines of evidence have implicated the short arm of chromosome 8 as harboring genes important in prostate carcinogenesis. Although most of this evidence comes from the identification of frequent somatic alterations of 8p loci in prostate cancer cells (e.g., loss of heterozygosity), studies have also suggested a role for 8p genes in mediation of inherited susceptibility to prostate cancer. To further examine this latter possibility, we performed linkage analyses, in 159 pedigrees affected by hereditary prostate cancer (HPC), using 24 markers on the short arm of chromosome 8. In the complete set of families, evidence for prostate cancer linkage was found at 8p22-23, with a peak HLOD of 1.84 ($P = .004$), and an estimate of the proportion of families linked (α) of 0.14, at D8S1130. In the 79 families with average age at diagnosis >65 years, an allele-sharing LOD score of 2.64 ($P = .0005$) was observed, and six markers spanning a distance of 10 cM had LOD scores >2.0. Interestingly, the small number of Ashkenazi Jewish pedigrees ($n = 11$) analyzed in this study contributed disproportionately to this linkage. Mutation screening in HPC probands and association analyses in case subjects (a group that includes HPC probands and unrelated case subjects) and unaffected control subjects were carried out for the putative prostate cancer–susceptibility gene, *PG1*, previously localized to the 8p22-23 region. No statistical differences in the allele, genotype, or haplotype frequencies of the SNPs or other sequence variants in the PG1 gene were observed between case and control subjects. However, case subjects demonstrated a trend toward higher homozygous rates of less-frequent alleles in all three *PG1* SNPs, and overtransmission of a *PG1* variant to case subjects was observed. In summary, these results provide evidence for the existence of a prostate cancer–susceptibility gene at 8p22-23. Evaluation of the *PG1* gene and other candidate genes in this area appears warranted.

Introduction

The short arm of chromosome 8, specifically 8p22-23, may harbor a prostate cancer–susceptibility gene(s) for the following reasons. First, multiple loci on 8p are the sites of frequent loss of heterozygosity (LOH) in a variety of human cancers, including prostate (Macoska et al. 1995; Bova et al. 1996; MacGrogan et al. 1996; Vocke et al. 1996; Deubler et al. 1997; Prasad et al. 1998), colon (Cunningham et al. 1993), breast (Chuaqui et al. 1995), ovarian (Cliby et al. 1993), liver (Emi et al. 1992), lung (Wistuba et al. 1999), bladder (Knowles et al. 1993), and head and neck cancer (Ransom et al. 1996). In prostate cancer, LOH for markers on 8p was found to be one of the most frequent somatic alterations, occurring in >60% of prostate cancers (Cunningham et al. 1996), and multiple homozygous deletions have been mapped to this chromosomal arm (Bova et al. 1996; Prasad et al. 1998). In addition, alterations of cancer-related genes in the region, such as *LZTS1*, have been identified in prostate cancer specimens and cell lines (Ishii et al. 1999).

Second, genomewide scans for prostate cancer–susceptibility genes in pedigrees affected with hereditary prostate cancer (HPC) have provided some evidence for prostate cancer linkage on 8p (Smith et al. 1996; Gibbs et al. 2000). In the 66 pedigrees affected by HPC ascertained by our group (Smith et al. 1996), there were positive linkage scores at 8p, with a two-point parametric LOD of 0.7 at D8S550, a multipoint LOD assuming heterogeneity (HLOD) of 0.81 ($P = .05$) and a multipoint nonparametric linkage score (NPL) of 2.02 ($P = .02$). Similarly, Gibbs et al. (2000) reported evidence for linkage at the marker D8S1106, ~5 cM from

Received March 8, 2001; accepted for publication May 15, 2001; electronically published July 6, 2001.

Address for correspondence and reprints: Dr. William B. Isaacs, Marburg 115, Johns Hopkins Hospital, 600 N. Wolfe Street, Baltimore, MD 21287. E-mail: wisaacs@jhmi.edu

0002-9297/2001/6902-0010$02.00

the marker D8S550. The maximum multipoint NPL score was 2.02 in 44 pedigrees with late age at onset (≥66 years).

Third, a candidate prostate cancer–susceptibility gene located at 8p22-23, *PG1*, was cloned by a haplotype-based association study conducted by Geneset (Cohen et al. 1999). In their study, a high-density array of biallelic markers, around D8S262 and D8S277 in the 8p23 region, was used to build haplotypes in case and control samples. By comparing 281 prostate cancer case subjects with 130 unaffected control subjects ascertained in France, they found significant differences in allele, genotype, and haplotype frequencies of several SNPs in the *PG1* gene between case and control subjects. The allele frequencies of G of SNP 477, T of SNP 99217, and A of SNP 467 in case subjects (control subjects) were 0.33 (0.24), 0.31 (0.23), and 0.26 (0.16), respectively, in their study. In their study, the haplotype frequencies of G-T-A for the three SNPs were 0.25 and 0.13 in case and control subjects, respectively, with an odds ratio (OR) of 2.17 ($P = .0002$). A single protein sequence, designated as the *PG1* gene, was identified in this candidate region. The function of this gene is unknown, and no follow-up studies have been presented.

We have three major objectives in the current study: first, evaluate evidence for linkage at 8p22-23, using densely spaced markers in 159 HPC families ascertained at Johns Hopkins Hospital; second, evaluate evidence for association in the *PG1* region using both the family-based approach in the 159 HPC families and the case-control approach in 249 case subjects with sporadic prostate cancer and 211 unaffected male control subjects; and third, screen the *PG1* gene for segregating mutations, using the single-strand conformation polymorphism (SSCP) method.

Methods

Family Collection

All 159 families with HPC were collected and studied at the Brady Urology Institute at Johns Hopkins Hospital in Baltimore. Families were ascertained from three resources. A majority of them were ascertained through referrals generated as a response to a letter by one of us (P.C.W.) to 8,000 urologists throughout the country. The second source was identified from family history records of the patient population seen at Johns Hopkins Hospital for treatment of prostate cancer. The remaining families came from the respondents to articles published in a variety of lay publications describing our studies of families affected with prostate cancer. Prostate cancer diagnosis was verified by medical records for each affected man studied. Age at diagnosis of prostate cancer was confirmed either through medical records or from two other independent sources. The mean age at diagnosis was 64.3 years for the case subjects in these families. Of the families, 84% are non-Jewish whites, 6.9% are Ashkenazi Jews, and 8.8% are black.

All 249 unrelated case subjects were recruited from among patients who underwent treatment for prostate cancer at the John Hopkins Hospital. The diagnosis of prostate cancer for all these subjects was confirmed by pathology reports. Preoperative prostate-specific antigen (PSA) levels, Gleason score, and pathological stages were available for 92, 244, and 245 of the 249 case subjects, respectively. Mean age at diagnosis for these case subjects was 58.6 years (range 37–73 years, SD 6.85). Family-history information was not obtained. Over 93% of the case subjects are white, and 3.2% are black.

From among men participating in screening programs for prostate cancer, 222 control subjects not affected with prostate cancer were selected. By applying the exclusion criteria of abnormal digital rectal examination (DRE) and abnormal PSA level (i.e., ≥4 ng/ml), 211 were eligible for the study. The mean age at examination was 58 years (range 40–80 years, SD 8.01). Of the eligible control subjects, >86% are white, and 7.1% are black. On the basis of interviews of eligible control subjects, 5.6% have a brother or father affected with prostate cancer.

Marker Genotyping

Twenty-one microsatellite markers spanning ~35 cM at 8p22-23 were genotyped in 159 families with HPC. These markers were selected from Marshfield comprehensive human genetic maps (Broman et al. 1998). Multiplex PCR, using fluorescently labeled primers (either fam, hex, or ned), was performed, and the resulting PCR fragments were separated using capillary electrophoresis performed with an ABI 3700 sequencer. The genotypes were scored using ABI software (GENOTYPER). A modified version of the program Linkage Designer was used to bin the alleles and check inheritance. The output from Linkage Designer was then analyzed further for any inconsistencies by running the LINKAGE software (Lathrop et al. 1984; Cottingham et al. 1993) without disease phenotype information. Marker allele frequencies were estimated from the 214 independent individuals in the data set (among them, 13 are Ashkenazi Jews and 19 are black). The marker order and distances estimated from the data using CRIMAP (Lander and Green 1987) were similar to the results in the Marshfield database. Thus, the intermarker distances of the Marshfield database were used in the analyses.

Three SNPs in the *PG1* gene were genotyped in all 159 HPC pedigrees, in the 249 unrelated case subjects affected with prostate cancer, and in 211 unaffected control subjects. All information (e.g. sequence, nomenclature, and designation of SNPs) for *PG1* was obtained from Cohen et al. (1999). SNP 477 (C→ G) is in intron

3, SNP 99217 (C→T) is in intron 5, and SNP 467 (G→C) is in the 3′ untranslated region. Marker D8S561 is an intragenic marker. Direct sequencing of PCR products was used to genotype the three SNPs. All the PCRs were performed in a 10-μl volume consisting of 30 ng genomic DNA, 0.2 μM each primer, 0.2 mM each dNTP, 1.5 mM $MgCl_2$, 20 mM Tris-HCl, 50 mM KCl, and 0.5 U *Taq* polymerase (Life Technologies). The primers for the SNP 477 were 5′-TGTTGATTTACAGGCGGC-3′ and 5′-GGAAAGGTACTCATTCATAG-3′. The primers for the SNP 99217 were 5′-GGTGGGAATTTACTATATG-3′ and 5′-GTTTATTTTGTGTGAGCTTTG-3′. The primers for the SNP 467 were 5′-AAGTTCACCTTCTCA-AGC-3′ and 5′-TGAAAGAGTTTATTCTCTGG-3′ (Cohen et al. 1999). These primers amplified 429-bp, 430-bp, and 420-bp fragments for SNP 477, SNP 99217, and SNP 467, respectively. PCR cycling conditions were as follows: 94°C for 4 min; followed by 28 cycles of 94°C for 30 s, 55°C for 30 s, and 72°C for 20 s with a final extension of 72°C for 2 min, except annealing temperature for SNP 477 was 60°C. All PCR products were purified using QuickStep PCR purification kit (Edge BioSystems) to remove dNTPs and excess primers. All sequencing reactions were performed using dye-terminator chemistry (BigDye) and then were precipitated using 63% ± 5% ethanol. Samples were loaded onto an ABI 3700 DNA Analyzer after adding 7 μl of formamide.

Mutation Screening

Probands from 92 families affected by HPC were screened for sequence variations in the eight exons of *PG1* using SSCP analysis. All eight exons were screened using 10 primer sets (see table 1) based on intronic sequence, as described by Cohen et al. (1999). Primers for this analysis were chosen with a minimum distance of 4 bp between primer 3′ base and exon boundaries. Four different electrophoresis and gel conditions were used to maximize detection of sequence variations: mutation detection enhancement (MDE) at room temperature, MDE supplemented with 5% glycerol at room temperature, MDE at 4°C, and MDE supplemented with 5% glycerol at 4°C. SSCP gels were loaded immediately after completion of the PCR reactions incorporating ^{33}P dATP, then subjected to electrophoresis at 4 W for ≥16 h. Fragment detection was accomplished by autoradiography. Abnormally migrating products were directly sequenced as described above. Exons containing sequence variations in HPC probands were analyzed in control individuals as well.

Statistical Analyses

Tests for Hardy-Weinberg equilibrium (HWE) for all the markers and for linkage disequilibrium (LD) between all pairs of markers were performed using independent individuals (pedigree founders and spouses of family members) of families with HPC and all sporadic case subjects and control subjects not affected with prostate cancer (computer program GDA; Weir et al. 1996). The HWE tests were based on exact tests, where a large number of the possible arrays was generated by permuting the alleles among genotypes and the proportion of these permuted genotypic arrays that have a smaller conditional probability than the original data were calculated. The LD tests were based on an exact test assuming multinominal probability of the multilocus genotype, conditional on the single-locus genotype (Zaykin et al. 1995). A Monte Carlo simulation was used to assess the significance, by permuting the single-locus genotypes among individuals in the sample to simulate the null distribution. The empirical *P* values of both the HWE and LD tests were based on 10,000 replicate samples.

Multipoint linkage analyses were performed using both parametric and nonparametric methods, implemented by the computer program GENEHUNTER-PLUS (Kruglyak et al. 1996; Kong and Cox 1997). For the parametric analysis, the same autosomal dominant model that was used by Smith et al. (1996) was assumed. Under this model, disease-gene frequency of .003, incomplete penetrance, and phenocopies were assumed. Specifically, affected men were assumed to be disease-gene carriers, with a fixed phenocopy rate of 15%, whereas all unaffected men aged <75 years and all women were assumed to be of unknown phenotype. In men aged ≥75 years, the lifetime penetrance of gene carriers was estimated to be 63%, and the lifetime risk of prostate cancer for noncarriers was 16% in this age class. Linkage in the presence of heterogeneity was as-

Table 1

Primers Used for Mutation Screening of *PG1* Exons

Exon	Primer	Annealing Temperature (°C)
1a-F	GCCGAGCTGAGAAGATGCTG	62
1a-R	CGGGAGCTCGGGTGGACGCC	
1c-F	CGCTGCCGCCGAGCTGAG	63
1c-R	GGCTCACCTGGACCCCGG	
2-F	CAACATCATTCGTCAGTTTC	57
2-R	ACCTAGGTTTCATGCAAATG	
3-F	CTGTGAAGAGCCTCATGTAC	62
3-R	AGAGAGAAAAGCATGGAAAC	
4-F	CTGGCCAATTGTTATTTTAA	53
4-R	AATTTAGAAACTGAGAGCTG	
5-F	ACCAAATTTGCTCTATGTCC	60
5-R	AAAGTATCTTTTCCAGGAAG	
6-F	TTAATGACGGCACTGATTG	53
6-R	AGGTGCGTGAACACACTTAC	
7-F	CTTTATATGACCATGAGTTC	46
7-R	CTGGAACTGTTGTTACTCAC	
8a-F	CAGCGTGTAATAGCTACCTG	62
8a-R	CACATACAGCTTCCTTCCAG	
8c-F	CCATCAATGTTGATCTTAAGTGG	50
8c-R	AATGTAGCACATCCCACTGTCTG	

sessed by use of Smith's admixture test for heterogeneity (Ott 1998). In this test, two types of families are assumed, one type linked to the disease locus with a proportion of α, and the other type is not linked, with the proportion $1-\alpha$. A maximum-likelihood approach was used to estimate α by maximization of the admixed LOD score (HLOD).

For the nonparametric analysis, the estimated marker identical-by-descent (IBD) sharing of alleles for the various affected relative pairs was compared with its expected values under the null hypothesis of no linkage. A statistic "Z_{all}" in the program was used (Whittemore and Halpern 1994). Allele-sharing LOD scores were then calculated, on the basis of Z_{all}, with equal weight assigned to all families, using the computer program ASM (Kong and Cox 1997).

Both HLOD and allele-sharing LOD can be converted to a χ^2 ($\chi^2 = 4.6 \times$ HLOD). Although the true distribution of the χ^2 under the null hypothesis of no linkage is unknown—especially in multipoint analysis—we assume that the distribution is a mixture of one that is degenerate at 0 and one that can be approximated by the distribution of the maximum of two independent χ^2 variables, each with 1 df (Faraway 1993). P values were thus calculated by $0.5 \times [1 - (1 - P_1)(1 - P_1)]$, where P_1 is the P value of χ^2 with 1 df.

Family-based association tests were performed for all six markers in the 159 families affected with HPC, using a software package FBAT (Laird et al. 2000). Unlike the classic transmission/disequilibrium test (TDT), which is limited to a specific pedigree structure (one genotyped proband and two genotyped parents per pedigree), the FBAT utilizes data from nuclear families, sibships, or a combination of the two to test for linkage and linkage disequilibrium between traits and genotypes. The test for linkage is valid when multiple affected members per pedigree are used, and the power to detect linkage is increased if there is an association. The test for association is valid if one affected member per pedigree is used (the genotypes of all the affected members can be included) or if the empirical variance is used to account for correlation between transmissions in families when linkage is present. In brief, the FBAT determines, from the data, an S statistic that is the linear combination of offspring genotypes and phenotypes. The distribution of the S statistic is generated by treating the offspring genotype data as random and conditioning on the phenotypes and parental genotypes. When the marker is biallelic, a Z statistic and its corresponding P value is calculated. When the marker is multiallelic, a χ^2 test is performed, with number of df equal to the number of alleles.

Population-based association tests were performed for the two polymorphisms in case and control subjects. An unconditional logistic regression was used to test for association between genotypes and affection status, adjusting for potential confounders such as age. The association tests were also performed for whites only, to decrease potential population stratification. The reported P values were not adjusted for multiple testing.

Haplotype frequencies in unrelated individuals were estimated for the three SNPs by maximum-likelihood estimation, using the best state of haplotype composition (see The Haplotype Estimation Help Page). The assumption of equal prior probabilities was made as a starting point for the expectation maximization (EM) algorithm.

Results

Linkage Results at 8p22-23 in 159 Pedigrees Affected by HPC

Both parametric and nonparametric multipoint linkage analyses provided evidence for linkage between a prostate cancer–susceptibility locus and markers on chromosome 8p in the complete 159 HPC pedigrees (fig. 1). The highest parametric HLOD was 1.84 (P = .004) with $\alpha = 0.14$, observed at D8S1130, 22 cM from 8pter at 8p22. HLOD scores $\geq$0.5 extended across ~22 cM, flanked by markers D8S1819 at 10 cM and D8S1135 at 32 cM from 8pter. The number of pedigrees that had LOD scores >0.3, >0.5, and >1 in the 22 cM region were 66, 33, and 4, respectively. In the first 66 pedigrees that were included in our previous genome-wide screen (Smith et al. 1996), the highest HLOD increased from 0.7 at D8S550 (21 cM) to 1.67 (P = .005; $\alpha = 0.24$) at D8S1130 (22 cM), because of the inclusion of fine-mapping markers. The 93 new pedigrees also provided evidence for linkage, with the highest HLOD of 0.77 (P = .06; $\alpha = 0.12$) at D8S552 (26 cM). For the nonparametric analyses, the highest allele-sharing LOD was 1.66 (P = .006) observed at D8S503, ~16 cM from 8pter in the complete family set. The highest allele-sharing LODs were 1.99 (P = .002) at D8S1130 and 0.34 (P = .21) at D8S552, respectively, in the first 66 and new 93 pedigrees.

Linkage analyses stratified by pedigree characteristics show that the pedigrees linked to 8p tend to have late onset, larger numbers of affected family members, and male-to-male disease transmission. Since the results from parametric and nonparametric were similar, only the results from nonparametric analyses are presented (table 2). The peak allele-sharing LOD was 2.64 (P = .0005) at D8S503 in the 79 pedigrees with mean age at onset $\geq$65 years, 1.41 (P = .01) at D8S503 in the 90 pedigrees with five or more affected family members, and 1.31 (P = .01) in the 99 pedigrees with male-to-male disease transmission. Evidence for linkage in this region was observed in non-Jewish white pedigrees (n = 133) and

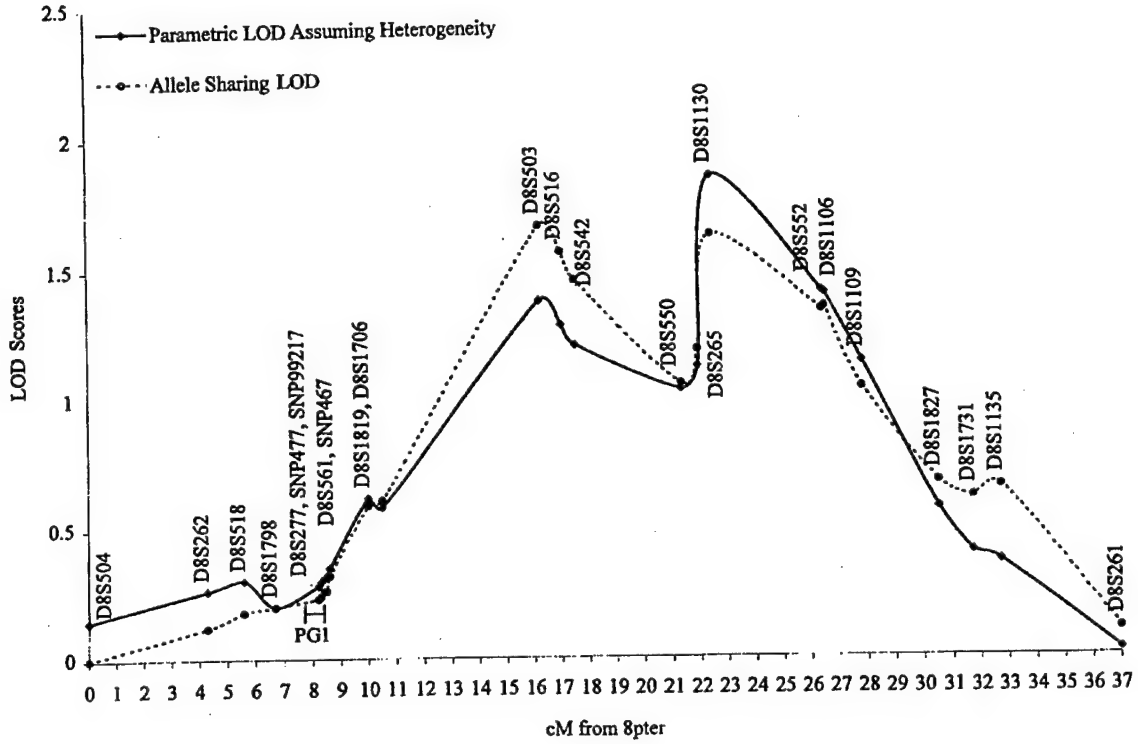


Figure 1 Results of multipoint parametric and nonparametric linkage analyses of prostate cancer–susceptibility loci, using 24 markers (21 microsatellite markers and 3 SNPs) on chromosome 8p22-23 in 159 families affected by HPC. The solid line represents parametric LOD under the assumption of heterogeneity. The dotted line represents allele-sharing LOD. Each diamond and circle represents a marker.

in the 11 Ashkenazi Jewish pedigrees (2 from the first 66 families), but not in the 14 black pedigrees. It is worth noting that 7 of the 11 Ashkenazi Jewish pedigrees had LOD scores ≥ 0.3 in the region and that, as a group, the 11 Ashkenazi families contributed disproportionately to the overall LOD score (table 2). By combining the non-Jewish white pedigrees with Ashkenazi pedigrees, we observed a LOD of 1.99 ($P = .002$) in the region.

To evaluate the impact of the marker allele frequencies on our linkage results in the black and Ashkenazi Jewish families, we repeated linkage analyses for the 14 black and 11 Ashkenazi families using marker allele frequencies estimated from 19 unrelated blacks and 13 unrelated Ashkenazi Jews, respectively. The results were similar to that using marker allele frequencies estimated from the mixed 214 unrelated subjects. In the 14 black families, the peak HLOD changed from 0.26 to 0.1 at D8S261. In the 11 Ashkenazi families, the peak HLOD changed from 1.25 to 1.24 at D8S1135. The robustness of our linkage results to the estimates of marker allele frequencies is probably due to the use of dense markers in multipoint analyses.

The evidence for linkage in and around the *PG1* gene (8 cM from pter) was weak. The highest HLOD and allele sharing LOD was 0.35 ($P = .18$) and 0.32 ($P = .20$), respectively, in the five markers within and surrounding the gene (from D8S277 to SNP 467).

Analysis of PG1: *Family-Based Linkage and Association Tests in 159 HPC Pedigrees*

Tests for HWE were performed for all microsatellite markers and SNPs analyzed, using 214 unrelated individuals from the 159 HPC pedigrees for which genotype information was available. All the markers tested were in HWE ($P > .05$). Marker-marker LD was tested for the five closely spaced markers (SNPs) in the *PG1* region. Markers SNP 477, SNP 99217, D8S561, and SNP 467 were in strong LD, with $P < .0001$ for all pairwise tests. Marker D8S277 was not in LD with these four markers (SNPs).

Family-based linkage and association tests were performed for the three SNPs. There was overtransmission of allele T of SNP 99217 from parents to affected sons, with $Z = 2.19$ ($P = .03$). The observed score S was 151 for allele T, compared with the expected 139. Similar

Am. J. Hum. Genet. 69:341–350, 2001

Table 2

Nonparametric Allele-Sharing LOD

		Nonparametric Allele-Sharing LOD								
		Age at Onset		No. of Affected Subjects		Male-to-Male Transmission		Ethnicity		
Markers	cM[a]	<65 (n = 79)	≥65 (n = 80)	<5 (n = 69)	≥5 (n = 90)	Yes (n = 99)	No (n = 60)	Black (n = 14)	Ashkenazi (n = 11)	Non-Jewish White (n = 133)
D8S504	0	0	.44	.00	.11	.2	0	0	.67	0
D8S262	4.3	0	.67	.00	.36	.11	.02	0	.78	.13
D8S518	5.6	0	.75	.00	.43	.15	.03	0	.93	.18
D8S1798	6.7	0	.91	.00	.43	.17	.04	0	1.01	.16
D8S277	8.2	0	1.01	.00	.52	.14	.09	0	.92	.18
SNP 477	8.3	0	1	.00	.55	.16	.09	0	.92	.19
SNP 99217	8.4	0	1.02	.00	.59	.18	.09	0	.92	.2
D8S561	8.5	0	.99	.00	.58	.17	.09	0	.92	.2
SNP 467	8.6	0	1.13	.00	.6	.24	.09	0	.92	.2
D8S1819	10	0	1.58	.00	.89	.47	.14	0	.93	.45
D8S1706	10.5	0	1.64	.00	.88	.58	.09	0	.96	.47
D8S503	16.2	0	2.64	.31	1.41	1.31	.39	0	1.27	1.38
D8S516	17	0	2.64	.24	1.41	1.16	.42	0	1.3	1.28
D8S542	17.5	0	2.61	.19	1.38	.98	.48	0	1.31	1.18
D8S550	21.3	0	1.96	.12	1.03	.69	.35	0	1.3	.81
D8S265	21.9	0	1.89	.13	1.17	.85	.34	0	1.26	.91
D8S1130	22.4	.1	1.97	.15	1.67	1.07	.56	0	1.12	1.39
D8S552	26.4	0	2.32	.28	1.09	.79	.54	0	.92	1.11
D8S1106	26.5	0	2.32	.28	1.09	.8	.54	0	.93	1.12
D8S1109	27.8	0	1.97	.11	1.03	.71	.33	0	1.07	.82
D8S1827	30.5	0	1.56	.01	.9	.84	.02	0	1.27	.36
D8S1731	31.7	0	1.55	.00	1.09	.9	0	0	1.51	.28
D8S1135	32.7	0	1.35	.00	1.13	.8	.02	0	1.6	.31
D8S261	37	0	.62	.00	.41	.11	.01	0	.85	.02

[a] Based on the Marshfield map.

tests for SNP 477 and SNP 467 were not significant, with $Z = 0.85$ ($P = .40$) and $Z = 0.31$ ($P = .76$), respectively. To decrease the impact of heterogeneity among races, the family-based linkage and association tests were performed again in the 133 non-Jewish white pedigrees. The test for SNP 99217 was significant, with $Z = 2.70$ ($P = .007$). The tests for the other two SNPs were not significant.

As either linkage or association in the data may lead to the significant test statistics, we performed two additional analyses to further explore the finding. The first analysis was a family-based association test using the empirical variance to account for correlation between transmissions in families when linkage is present. In this analysis, the evidence for association decreased, with $Z = 1.66$ ($P = .10$) and $Z = 2.07$ ($P = .04$), respectively in the complete 159 HPC pedigrees and in 133 non-Jewish white pedigrees. The second analysis is the stratified linkage analyses based on the probands' genotype at SNP 99217. The pedigrees whose probands are T carriers contributed disproportionally to the evidence for linkage at 5 markers in the region. The 77 pedigrees whose probands are heterozygous 'T' and the 15 pedigrees whose probands are homozygous 'T' carriers had allele-sharing LODs of 0.5 ($P = .12$) and 1.44 ($P = .01$) at SNP 99217, respectively. In contrast, the 78 pedigrees whose probands are not T carriers had HLOD of 0. These data suggest that both linkage and association contribute to the significance of the family-based test.

Analysis of PG1: *Population-Based Association Tests in HPC Probands, Unrelated Case Subjects, and Unaffected Control Subjects*

The three *PG1* SNPs were genotyped in all 159 HPC pedigrees and in 249 unrelated prostate cancer case subjects and 211 unaffected control subjects. All SNPs were in HWE in each subset. Allele frequencies of the three SNPs were compared between case and control subjects. To decrease the confounding factor of racial differences, the comparison was limited to whites only. For SNP 477, the allele frequencies of G were 0.33, 0.33, and 0.31, in the 123 HPC probands, 216 unrelated case subjects, and

178 unaffected control subjects, respectively. For SNP 99217, the allele frequencies of T were 0.32, 0.31, and 0.30, in the 131 HPC probands, 222 unrelated case subjects, and 177 unaffected control subjects, respectively. For SNP 467, the allele frequencies of A were 0.24, 0.25, and 0.24, in the 120 HPC probands, 210 unrelated case subjects, and 177 unaffected control subjects, respectively. No significant difference was observed in the allele frequencies between the probands and control subjects, between the unrelated case subjects and control subjects, or between all case subjects and control subjects in any of the three SNPs.

Genotype frequencies of the three SNPs were also compared in the white subjects only (table 3). No statistical differences in genotype frequencies were observed between case and control subjects for any of the three SNPs. There was a trend toward higher homozygous rates of the less-frequent alleles of each SNP in the case subjects with HPC and in the unrelated case subjects, compared with those in the control subjects; however, the differences were not statistically significant. For example, the odds ratio (OR) was 1.39 (95% confidence interval [CI] 0.73–2.63) when the homozygous frequencies for T/T of SNP 477 in all case and control subjects were compared.

Haplotype frequencies of the three SNPs were also compared between case and control subjects. The estimated haplotype frequencies of G-T-A for the three SNPs (SNP 477, SNP 99217, and SNP 467) were 0.21, 0.25, and 0.22, in HPC probands, unrelated case subjects, and unaffected control subjects, respectively. No significant statistical differences in the haplotype frequencies were found between all possible pair comparisons.

Mutation Screening of PG1

SSCP mutation-screening analysis of probands from 92 families with HPC produced band patterns indicative of two different sequence variants in exon 1 and three different variants in exon 4. For exon 1, sequence analysis identified one variant as a silent polymorphism (C→G at codon 43, position 2159 in the genomic sequence reported by Cohen et al. [1999]), which was present in 14.1% of probands and in 7.9% of unaffected control subjects. The other variant was a nonsynonymous change at codon 22 (G→C at position 2095, resulting in substituting Ala for Gly), present in 4.2% of probands and 2.2% of control subjects.

Sequence analysis of the variants in exon 4 demonstrated two silent polymorphisms (T→C in codon 145 at position 25631 in the genomic sequence reported by Cohen et al. [1999], and A→G in codon 139 at position 25615) and a nonsynonymous change at position 25649 (G→A resulting in a substitution of Thr for Ala at codon 151). These variants were present at low frequencies (0.5%–3%) with no differences between case and control subjects (e.g., the Ala→Thr change was observed in one proband, one sporadic case subject, and one control individual).

Discussion

By testing for linkage and association between prostate cancer susceptibility and markers on 8p22-23 in 159 HPC pedigrees, 249 unrelated case subjects, and 211 unaffected control subjects, we obtained the following three findings. (1) There was evidence for linkage be-

Table 3

Genotypes of Three SNPs in *PG1* in Probands, Unrelated Case Subjects, and Unaffected Control Subjects (White Subjects Only)

SNP and Genotype	Control Subjects (%)	Case Subjects (%)		Odds Ratio* (95% CI)		
		Sporadic	HPC	Sporadic Case Subjects vs. Control Subjects	HPC Case Subjects vs. Control Subjects	All Case Subjects vs. Control Subjects
SNP 477:	$n=178$	$n=222$	$n=123$			
C/C	.47	.46	.46	1	1	1
C/G	.44	.41	.42	1.06 (.70–1.60)	.96 (.59–1.58)	1.03 (.70–1.51)
G/G	.09	.13	.12	1.38 (.69–2.74)	1.36 (.61–3.04)	1.39 (.73–2.63)
Any G				1.11 (.75–1.65)	1.03 (.64–1.64)	1.09 (.76–1.57)
SNP 99217:	$n=177$	$n=217$	$n=131$			
C/C	.49	.51	.47	1	1	1
C/T	.42	.36	.41	.97 (.64–1.48)	.99 (.61–1.61)	.99 (.67–1.44)
T/T	.08	.13	.11	1.20 (.59–2.45)	1.33 (.60–2.97)	1.25 (.65–2.41)
Any T				1.01 (.68–1.51)	1.05 (.66–1.67)	1.03 (.72–1.48)
SNP 467:	$n=177$	$n=212$	$n=120$			
G/G	.59	.59	.60	1	1	1
G/A	.34	.32	.32	1.10 (.72–1.69)	.92 (.55–1.54)	1.03 (.69–1.53)
A/A	.07	.10	.08	1.30 (.60–2.84)	1.20 (.49–2.93)	1.28 (.63–2.63)
Any A				1.14 (.76–1.71)	.97 (.60–1.57)	1.08 (.74–1.56)

* All odds ratios were adjusted for age.

tween a prostate cancer–susceptibility locus and markers on 8p22-23, with a highest HLOD of 1.84 ($P = .004$) at D8S1130. The region providing evidence for linkage spanned ~22 cM at 8p22-23. The evidence for linkage was observed in the first 66 HPC pedigrees and in the 93 new HPC pedigrees. The pedigrees with late age at onset, a large number of affected family members, and male-to-male disease transmission provided stronger evidence for linkage at the region. (2) One intronic sequence variant (allele T of SNP 99217) in the putative prostate cancer–susceptibility gene (*PG1*) was overtransmitted from parents to affected offspring, with $Z = 2.19$ ($P = .03$) and $Z = 2.70$ ($P = .007$) in all 159 HPC pedigrees and in 133 non-Jewish white pedigrees, respectively. The overtransmission of allele T likely reflected evidence for both linkage and association in the data, since (*a*) a family-based association test that accounted for the presence of linkage provided weaker but still marginally significant test statistics, with $Z = 1.66$ ($P = .10$) and 2.07 ($P = .04$) in all HPC pedigrees and in non-Jewish white pedigrees, and (*b*) families whose probands carry T are more likely to be linked to the *PG1* gene region. (3) No statistical differences were found in the allele, genotype, and haplotype frequencies for the three SNPs or other sequence variants in the *PG1* gene between HPC probands, unrelated prostate cancer case subjects, and unaffected control subjects. However, a trend (but not a statistically significant one) was observed toward higher homozygous rates of the less-frequent allele of each SNP in the HPC case subjects and in the unrelated case subjects, compared with those among the control subjects.

Evidence for linkage at 8p22-23 in our study did not reach the genomewide screen criteria for significant or suggestive linkage as proposed by Lander and Kruglyak (1995). However, we think our results provide a basis for further study in this region for a number of reasons. First, the prior probability that a prostate cancer–susceptibility gene lies near 8p22-23 is high as extensive evidence from LOH studies in prostate and other cancers indicates the existence of tumor-suppressor genes in the region (for review, see work by Bookstein [2001]). Therefore, the stringent criterion for significant linkage, which is used to account for the low prior probability of any pair of genes being located within a recombination fraction of $<.5$ in the human genome, is not appropriate in this situation (Ott 1998). Secondly, although the HLOD of 1.84 ($P = .004$) could represent false-positive evidence for linkage, our simulation results suggested that it is unlikely. On the basis of the same structure of 159 pedigrees with HPC (affection status and availability of genotyping) and the genetic model used in the analyses, we simulated 10,000 replicates with a six-allele marker (equally frequent) not linked to the disease gene using FASTSLINK (see D. Weeks's FTP page). We then analyzed each replicate and only observed 10 of the 10,000 replicates with a HLOD >1.84, yielding an empirical P value of .001. Thirdly, and perhaps most importantly, the same region was reported to be linked to a prostate cancer–susceptibility gene in an independent genomewide-screen linkage study. Gibbs et al. (2000) reported a maximum multipoint nonparametric linkage score of 2.02 at D8S1106 in 44 pedigrees with late age at onset (≥66 years), using genomewide screen markers. This marker was in our linkage region, ~5 cM from the peak marker, D8S1130. Interestingly, we observed the same trend that pedigrees with late age at onset tend to be linked to this region, with a peak allele sharing LOD of 2.64 ($P = .0005$) in our 80 pedigrees with age at onset ≥65 years. Lastly, both series of our HPC pedigrees (the first 66 HPC pedigrees included in the initial genomewide screen and the 93 pedigrees ascertained later) provided evidence for linkage. The trend for this linkage to be more prominent in families with older age at diagnosis was observed in both the first and the second groups of families (allele-sharing LOD scores of 1.46, $P = .009$ and 1.32, $P = .01$ respectively).

Even though some evidence for linkage at the *PG1* gene was observed in parametric and nonparametric linkage analyses and family-based linkage and association test, the rather weak linkage at *PG1* gene and the distance (10–15 cM) between the *PG1* gene and the highest linkage region indicated that the *PG1* plays a minor role, if any, in accounting for the linkage signal at 8p22-23. One or several other genes in the region may contribute to the observed linkage. Several important candidate tumor-suppressor genes reside in the 8p22-23 region, including the *N33* (Bova et al. 1996), macrophage-scavenger-receptor (*MSR*) (Kagan et al. 1995; Bova et al. 1996), the N-acetyltransferase genes *NAT1* and *NAT2* (Wang et al. 1999), *LZTS1* (Ishii et al. 1999), and *DLC1* (deleted in liver cancer; see Yuan et al. 1998; Wilson et al. 2000). Several mutations in *LZTS1* were found in prostate cancer cell lines. Transcript analysis from several *LZTS1*-expressing tumors revealed truncated mRNAs, including a frameshift (Ishii et al. 1999). Mutations in *DLC1* were found in colorectal and ovarian tumors (Wilson et al. 2000). Unfortunately, studies investigating possible associations between the genomic sequence variants and prostate cancer have not been published.

The interpretation of the results from our *PG1* gene–association study is difficult. Although overtransmission of allele T of SNP 99217 from parents to affected offspring provides evidence that *PG1* might influence prostate cancer susceptibility, the lack of statistically significant differences in the allele, genotype, and haplotype frequencies between case and control subjects is not consistent with this notion. Our re-

sults contrast with the results from the case-control study reported by Cohen et al. (1999). Although the exact reason for the difference is unknown, several of the following factors may contribute to the difference. First, there may be allele-frequency differences between the French and U.S. populations, and the former may be a more homogeneous population. This is, however, unlikely to be the major reason in this case, because the allele frequencies in the case subjects are similar in the two populations. Second, the power to detect the association in our study sample is limited. Using the point estimates of ORs and frequencies from Cohen et al. (1999), the power to detect an OR of 2.2, at the significance level of .05, with a genotype frequency of 14% in control subjects, is 72% in our combined 345 case and 177 control samples (white subjects only). Third, potential misclassification may be present in our control group. Although the unaffected control subjects in our study had normal results on digital rectal examination and normal PSA levels (i.e., <4 ng/ml), some of our control subjects are young, and they could be disease-gene carriers who will develop prostate cancer later. The ORs adjusted for age in our study may alleviate the problem but cannot remove the confounder. Last, random sampling error in control subjects in both studies could lead to the difference. In consideration of the limited power to detect a weak association and potential bias in the study, further studies utilizing larger number of control subjects may help to answer the question.

In summary, our study provides evidence for prostate cancer linkage at 8p22-23. The linkage results, along with the consistent evidence that 8p22 is the most commonly deleted region in prostate cancer cells and the discovery of mutations in some tumor-suppressor genes in the region warrant further studies. The results of the evaluation of the *PG1* gene are inconclusive but interesting enough to suggest further studies of this gene as well. With the availability of more-complete sequence data for the human genome, studies to systematically evaluate all the genes in the region using an association study design (either case-control or family-based) are justified and likely to succeed.

Acknowledgments

The authors thank all the study subjects who participated in this study. This work was partially supported by National Cancer Institute SPORE grant CA58236 and by two grants from the Department of Defense (to W.B.I. and J.X.).

Electronic-Database Information

Accession numbers and URLs for data in this article are as follows:

GDA: Software for the Analysis of Discrete Genetic Data, http://lewis.eeb.uconn.edu/lewishome/gda.html

Haplotype Information Help Page, http://www.bioinf.mdc-berlin.de/hap/ithap-help.html

Linkage Designer, http://dnalab-www.uia.ac.be/dnalab/ld.html

Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim/ (for prostate cancer [MIM 176807], HPC2 [MIM 605367], HPC1 [MIM 601518], PCaP [MIM 602759], PCBP/CAPB [MIM 603688], and HPCX [MIM 300147])

Weeks FTP page, ftp://watson.hgen.pitt.edu (for FASTSLINK)

References

Bova GS, MacGrogan D, Levy A, Pin SS, Bookstein R, Isaacs WB (1996) Physical mapping of chromosome 8p22 markers and their homozygous deletion in a metastatic prostate cancer. Genomics 35:46–54

Bookstein R (2001) Tumor suppressor genes in prostate cancer. In: Chung LWK, Isaacs WB, Simons JW (eds) Prostate cancer, biology, genetics, and the new therapeutics. Humana Press, Totowa, NJ

Broman KW, Murray JC, Sheffield VC, White RL, Weber JL (1998) Comprehensive human genetic maps: individual and sex-specific variation in recombination. Am J Hum Genet 63:861–689

Chuaqui RF, Sanz-Ortega J, Vocke C, Linehan WM, Sanz-Esponera J, Zhuang Z, Emmert-Buck MR, Merino MJ (1995) Loss of heterozygosity on the short arm of chromosome 8 in male breast carcinomas. Cancer Res 55: 4995–4998

Cliby W, Ritland S, Hartmann L, Dodson M, Halling KC, Keeney G, Podratz KC, Jenkins RB (1993) Human epithelial ovarian cancer allelotype. Cancer Res 53:2393–2398

Cohen D, Chumakov I, Blumenfeld M, Bougueleret L (1999) GENSET, assignee. Prostate cancer gene. US patent US005945522A. August 31, 1999

Cottingham RW Jr, Idury RM, Schaffer AA (1993) Faster sequential genetic linkage computations. Am J Hum Genet 53: 252–263

Cunningham C, Dunlop MG, Wyllie AH, Bird CC (1993) Deletion mapping in colorectal cancer of a putative tumour suppressor gene in 8p22-p21.3. Oncogene 8:1391–1396

Cunningham JM, Shan A, Wick MJ, McDonnell SK, Schaid DJ, Tester DJ, Qian J, Takahashi S, Jenkins RB, Bostwick DG, Thibodeau SN (1996) Allelic imbalance and microsatellite instability in prostatic adenocarcinoma. Cancer Res 1996 56:4475–4482

Deubler DA, Williams BJ, Zhu XL, Steele MR, Rohr LR, Jensen JC, Stephenson RA, Changus JE, Miller GJ, Becich MJ, Brothman AR (1997) Allelic loss detected on chromosomes 8, 10, and 17 by fluorescence in situ hybridization using single-copy P1 probes on isolated nuclei from paraffin-embedded prostate tumors. Am J Pathol 150:841–850

Emi M, Fujiwara Y, Nakajima T, Tsuchiya E, Tsuda H, Hirohashi S, Maeda Y, Tsuruta K, Miyaki M, Nakamura Y (1992) Frequent loss of heterozygosity for loci on chromosome 8p in hepatocellular carcinoma, colorectal cancer, and lung cancer. Cancer Res 52:5368–5372

Faraway JJ (1993) Distribution of the admixture test for the detection of linkage under heterogeneity. Genet Epidemiol 10:75–83

Gibbs M, Stanford JL, Jarvik GP, Janer M, Badzioch M, Peters MA, Goode EL, Kolb S, Chakrabarti L, Shook M, Basom R, Ostrander EA, Hood L (2000) A genomic scan of families with prostate cancer identifies multiple regions of interest. Am J Hum Genet 67:100–109

Ishii H, Baffa R, Numata SI, Murakumo Y, Rattan S, Inoue H, Mori M, Fidanza V, Alder H, Croce CM (1999) The FEZ1 gene at chromosome 8p22 encodes a leucine-zipper protein, and its expression is altered in multiple human tumors. Proc Natl Acad Sci USA 96:3928–3933

Kagan J, Stein J, Babaian RJ, Joe YS, Pisters LL, Glassman AB, von Eschenbach AC, Troncoso P (1995) Homozygous deletions at 8p22 and 8p21 in prostate cancer implicate these regions as the sites for candidate tumor suppressor genes. Oncogene 11:2121–2126

Knowles MA, Shaw ME, Proctor AJ (1993) Deletion mapping of chromosome 8 in cancers of the urinary bladder using restriction fragment length polymorphisms and microsatellite polymorphisms. Oncogene 8:1357–1364

Kong A, Cox NJ (1997) Allele-sharing models: LOD scores and accurate linkage tests. Am J Hum Genet 61:1179–1188

Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES (1996) Parametric and nonparametric linkage analysis: a unified multipoint approach. Am J Hum Genet 58:1347–1363

Laird NM, Horvath S, Xu X (2000) Implementing a unified approach to family-based test of association. Genet Epidemiol Suppl 19:36–42

Lander ES, Green P (1987) Construction of multilocus genetic linkage maps in humans. Proc Natl Acad Sci USA 84:2363–2367

Lander ES, Kruglyak L (1995) Genetic dissection of complex traits: guideline for interpreting and reporting linkage results. Nat Genet 11:241–247

Lathrop GM, Lalouel JM, Julier C, Ott J (1984) Strategies for multilocus linkage analysis in humans. Proc Natl Acad Sci USA 81:3443–3446

MacGrogan D, Levy A, Bova GS, Isaacs WB, Bookstein R (1996) Structure and methylation-associated silencing of a gene within a homozygously deleted region of human chromosome band 8p22. Genomics 35:55–65

Macoska JA, Trybus TM, Benson PD, Sakr WA, Grignon DJ, Wojno KD, Pietruk T, Powell IJ (1995) Evidence for three tumor suppressor gene loci on chromosome 8p in human prostate cancer. Cancer Res 15:5390–5395

Ott J (1998) Analysis of human genetic linkage. 3d ed. Johns Hopkins Press, Baltimore

Prasad MA, Trybus TM, Wojno KJ, Macoska JA (1998) Homozygous and frequent deletion of proximal 8p sequences in human prostate cancers: identification of a potential tumor suppressor gene site. Genes Chromosomes Cancer 23: 255–262

Ransom DT, Leonard JH, Kearsley JH, Turbett GR, Heel K, Sosars V, Hayward NK, Bishop JF (1996) Loss of heterozygosity studies in squamous cell carcinomas of the head and neck. Head Neck 18:248–253

Smith JR, Freiji D, Carpten JD, Gronberg H, Xu J, Isaacs SD, Brownstei MJ, Bova GS, Guo H, Bujnovsky P, Nusskern DR, Damber JE, Bergth A, Emanuelsson M, Kallioniemi OP, Walker-Daniels J, Baily-Wilson JE, Beaty TH, Meyers DA, Walsh PC, Collins FS, Trent JM, Isaacs WB (1996) Major susceptibility locus for prostate cancer on chromosome 1 suggested by a genome-wide search. Science 274:1371–1374

Vocke CD, Pozzatti RO, Bostwick DG, Florence CD, Jennings SB, Strup SE, Duray PH, Liotta LA, Emmert-Buck MR, Linehan WM (1996) Analysis of 99 microdissected prostate carcinomas reveals a high frequency of allelic loss on chromosome 8p12-21. Cancer Res 56:2411–2416

Wang CY, Debiec-Rychter M, Schut HA, Morse P, Jones RF, Archer C, King CM, Haas GP (1999) N-Acetyltransferase expression and DNA binding of N-hydroxyheterocyclic amines in human prostate epithelium. Carcinogenesis 20: 1591–1595

Weir BS (1996) Genetic data analysis II: methods for discrete population genetic data. Sinauer, Sunderland, MA

Whittemore A, Halpern J (1994) A class of tests for linkage using affected pedigree members. Biometrics 50:118–127

Wilson PJ, McGlinn E, Marsh A, Evans T, Arnold J, Wright K, Biden K, Young J, Wainwright B, Wicking C, Chenevix-Trench G (2000) Sequence variants of DLC1 in colorectal and ovarian tumours. Hum Mutat 15:156–165

Wistuba II, Behrens C, Virmani AK, Milchgrub S, Syed S, Lam S, Mackay B, Minna JD, Gazdar AF (1999) Allelic losses at chromosome 8p21-23 are early and frequent events in the pathogenesis of lung cancer. Cancer Res 59:1973–1979

Yuan BZ, Yang Y, Keck-Waggoner CL, Zimonjic DB, Thorgeirsson SS, Popescu NC (1999) Assignment and cloning of mouse Arhgap7 to chromosome 8A4-B2, a conserved syntenic region of human chromosome 8p22→p21. Cytogenet Cell Genet 87:189–190

Zaykin D, Zhivotovsky L, Weir BS (1995) Exact tests for association between alleles at arbitrary numbers of loci. Genetica 96:169–178

[CANCER RESEARCH 62, 2253–2257, April 15, 2002]

Advances in Brief

Associations between *hOGG1* Sequence Variants and Prostate Cancer Susceptibility[1]

Jianfeng Xu, Siqun L. Zheng, Aubrey Turner, Sarah D. Isaacs, Kathy E. Wiley, Gregory A. Hawkins, Bao-li Chang, Eugene R. Bleecker, Patrick C. Walsh, Deborah A. Meyers,[2] and William B. Isaacs

Center for Human Genomics, Wake Forest University School of Medicine, Winston-Salem, North Carolina 27157 [J. X., S. L. Z., A. T., G. A. H., B-l. C., E. R. B., D. A. M.], and Department of Urology, Johns Hopkins Medical Institutions, Baltimore, Maryland 21287 [S. D. I., K. E. W., P. C. W., W. B. I.]

Abstract

8-Hydroxyguanine is a mutagenic base lesion produced by reactive oxygen species. The *hOGG1* gene encodes a DNA glycosylase/AP lyase that can suppress the mutagenic effects of 8-hydroxyguanine by catalyzing its removal from oxidized DNA. A population-based (245 cases and 222 controls) and family-based (159 hereditary prostate cancer families) association study was performed to test the hypothesis that sequence variants of *hOGG1* increase susceptibility to prostate cancer. We found that the genotype frequency of two sequence variants (11657A/G and Ser326Cys) was significantly different between cases and controls. The association with 11657A/G is confirmed and strengthened by our family-based association study. These results suggest that sequence variants in this gene are associated with prostate cancer risk, presumably through defective DNA repair function of *hOGG1*.

Introduction

The DNA repair enzyme OGG1 is a DNA glycosylase/AP lyase that has been hypothesized to play an important role in preventing carcinogenesis by repairing oxidative damage to DNA (1). Specifically, glycosylase/AP lyase can efficiently repair 8-OH-G[3] a major base lesion produced by ROS, formed as a byproduct of endogenous metabolism or exposure to environmental oxidizing agents, such as ionizing radiation or chemical genotoxic compounds. 8-OH-G is highly mutagenic and, if not excised on DNA replication, can cause GC to TA transversions, which occur frequently in several oncogenes and tumor suppressor genes (2).

The genomic DNA of *hOGG1*, with eight exons, spans ~16.7 kb on 3p25. Several SNPs in the *hOGG1* gene have been identified, and the repair activities of the variant proteins have been evaluated in many studies (3–6). However, in contrast to these extensive functional studies, limited knowledge is available on the association between cancer susceptibility and SNPs in this critical DNA repair gene. To date, only five studies have been reported on the association between *hOGG1* SNPs and cancer susceptibility, and all of these have focused on a frequently observed missense change at codon 326 in exon 7 (Ser326Cys). Although three of these previous studies did not find statistical differences in the genotype distributions of the SNP between cancer cases and normal controls (3, 7–8), two studies found a significantly increased frequency of Cys/Cys in lung and esophageal cancer cases (9–10). Furthermore, a significant difference in the distribution of Ser326Cys was observed between ethnicities, with the frequency of Ser326 being 0.78 and 0.59 in Caucasian and Asian controls, respectively.

Although sequence variants in genes involved in DNA repair may be an important determinant of inherited susceptibility to cancer in humans (11), this could be particularly relevant for prostate cancer, in which oxidative damage has been proposed to play a critical role in cancer formation. Indeed, the preventative effect of antioxidants and the cancer-associated induction and molecular inactivation of components of the cellular defense system for oxidative stress have been cited as evidence of the important procarcinogenic aspect of ROS in the human prostate (12). In addition, the *hOGG1* gene is abundantly expressed in prostate tissue. Finally, a study by Osterod *et al.* (13) found that the accumulation of oxidative DNA base damage in *OGG1*-deficient mice is age related and tissue specific. Although we do not know whether this model is directly applicable to prostate, we can hypothesize that the accumulated effect of altered DNA repair activities associated with sequence variants has a larger impact on this late age of onset cancer.

On the basis of the present understanding of the *hOGG1* gene function in the DNA repair pathway and the existing epidemiological data, we hypothesized that sequence variants of the *hOGG1* gene confer risk to prostate cancer. Therefore, we tested the following four subhypotheses: (*a*) the missense change Ser325Cys is associated with increased risk to prostate cancer; (*b*) other sequence variants in the *hOGG1* gene are associated with prostate cancer risk; (*c*) sequence variants of *hOGG1* may produce a different risk to hereditary *versus* sporadic prostate cancer; and (*d*) clinical characteristics of sporadic prostate cancer are associated with sequence variants of *hOGG1*.

Subjects and Methods

Subjects. A detailed description of the study sample was presented previously (14). HPC families (n = 159) were ascertained at the Brady Urology Institute at Johns Hopkins Hospital (Baltimore, MD), through referrals, review of medical records for patients seen at Johns Hopkins Hospital for treatment of prostate cancer, and respondents to various lay publications describing our studies. Each family had at least three men affected with prostate cancer. The mean number of affected men per family was 5.1, and the mean age at diagnosis was 64.3 years. The majority of HPC families were Caucasians (n = 133; 84%), and there were 14 (8.8%) African-American families. For the 159 probands of these families, the mean age at diagnosis was 61 years. The diagnosis of prostate cancer was verified by medical records.

All of the 245 unrelated prostate cancer cases were recruited from patients who underwent treatment for prostate cancer at the Johns Hopkins Hospital and did not have first-degree relatives affected with prostate cancer. For each subject, the diagnosis of prostate cancer was confirmed by pathology reports. Preoperative PSA levels, Gleason score, and pathological stages were available for 202, 240, and 241 cases, respectively. Mean age at diagnosis for these cases was 58.7 years. More than 93% of the cases were Caucasian, and 3.2% were African American.

Two hundred twenty-two non-prostate cancer controls were selected from men participating in screening programs for prostate cancer. By applying the

Received 12/17/01; accepted 3/1/02.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[1] This work was partially supported by PHS SPORE CA58236 and two grants from the Department of Defense (to W. B. I. and J. X.).

[2] To whom requests for reprints should be addressed, at Center for Human Genomics, Wake Forest University School of Medicine, Medical Center Boulevard, Winston-Salem, NC 27157. Phone: (336) 716-5700; Fax: (336) 716-7575; E-mail: dmeyers@wfubmc.edu.

[3] The abbreviations used are: 8-OH-G, 8-hydroxyguanine; ROS, reactive oxygen species; SNP, single nucleotide polymorphism; HPC, hereditary prostate cancer; PSA, prostate-specific antigen; DRE, digital rectal examination; HWE, Hardy-Weinberg equilibrium; LD, linkage disequilibrium; FET, Fisher's exact test; CI, confidence interval; RR, relative risk.

Table 1 *PCR primers for the SNPs in hOGG1 gene*

SNP[a]	Group[b]	Primers[c]		
		Forward	Reverse	Extension (direction)
−627T/C	1	TGGTGAAGACAGGGTTCGTG	TTCTCCGAGAGCCGTTCTCC	CTCCCCCGAGCCTTTGCA (R)
−23A/G	1	GCATTTCCACAGCAGGCACC	AAGGGTCGTGGTCCTTGTCTG	CTGGGTAGGCGGGGCTACT (F)
−18G/T	5	TTGTCTGGGCGGGGTCTTTG	GGCAGGCATTTCCACAGCAG	ATTTCCACAGCAGGCACCG (R)
2550A/C	1	GAGGTCGAGGCAGGCAGAT	GGTTTCACCATGTTGGCCAG	CTGGTCTTGAACTCCTGACC (R)
3224A/C	3	CCCCGTCTCTACTAAAAATAC	TTCACGCCATTCTCCTGCCT	CGCCCACCACCACACCC (R)
3402G/A	2	ACAGAGTGAGACTCTGTCTC	GCCTTATGACTAACTAAGCC	TGACTAACTAAGCCAGGAGC (R)
3574G/A	2	GCAATCATGAGGCAGTGTAG	GCTGAAATTACCAGCATGAG	CAGCATGAGCTACCCCACC (R)
4540G/A	2	GAGCCATCCTGGAAGAACAG	CTTGTGGGCCTCCTCATATG	GGCCTCCTCATATGAGGACTCT (R)
6170G/C	3	GCTATAAGCAAGATGCTGGC	TGCAGTCAGCCACCTTTGAC	AGCCACCTTTGACAGACACA (R)
6803C/G		TGCCCAACGTGACTACAGC	GAGGTAGTCACAGGGAGGCC	
6876T/A	5	AGAGAAGTGGGGAATGGAGG	AGAAAGGGTTCCAAAGGGCC	AGGCTAGATGGGGCACCC (F)
6893T/C	4	AGAAAGGGTTCCAAAGGGCC	AGAGAAGTGGGGAATGGAGG	GAATGGAGGGGAAGGTGCTT (R)
7143A/G	3	CTGAACCGGGAGTTTCTCTG	GGAAAATGCAGTGAGGAGTG	ATGCAGTGAGGAGTGGTAGGGA (F)
9110A/G	4	TACCCAGGCTCAGCTTGCAC	AGTCCTTCAGTAAGGATCCC	AAGCAGTTACTGTGTGCCCA (F)
10629C/G	4	GGAGTTCCCCCTTTATAAAC	CAAGGAACAGAAAGGATAATG	CAGAAAGGATAATGTAGCTAGAA (R)
10660A/T	5	CTAGCTACATTATCCTTTCTG	AAGCAATGGCAAGTGCAAAG	GAGGGCAAGATGGCGGCACAT (R)
11657A/G	6	AGGTTTAGAGACAGTTCCCC	CAAGGAAGCTCTCAAGAAGG	CCAGGAAGGACAAGGCTCA (F)
11826A/T	7	AATGCCATCCTCACTGCTTC	AGTCACTTTGCCTCCAAAGG	TGCCTCCAAAGGCATCAGTT (R)

[a] Numerical values represent the position (measured in base pairs) from the transcription site. The letters represent nucleotide change.
[b] Multiplex PCR group.
[c] All have the ACGTTGGATG tag in the front, except for the SNP 6803C/G (by direct sequencing). R, reverse; F, forward.

exclusion criteria of abnormal DRE and abnormal PSA level (*i.e.*, ≥4 ng/ml), 211 were eligible for the study. The mean age at examination was 58 years. More than 86% of the eligible controls were Caucasian and 7.1% were African American. On the basis of interviews of the subjects, we learned that 5.6% of the eligible controls had brothers or their father affected with prostate cancer.

The Institutional Review Board of Johns Hopkins University approved the protocols for subject recruitment. After each participant was guided through an informed consent process, they completed and signed a consent form as a record of this process.

Sequencing Methods and SNP Genotyping. SNPs information was obtained from the Celera database. All of the SNPs, except one, were genotyped using the MassARRAY system (SEQUENOM, Inc., San Diego, CA). Table 1 lists the PCR primers and extension primers for all of the SNPs. SNP Ser326Cys was genotyped using direct sequencing. Sequence reaction was run in the ABI 3700 DNA analyzer and analyzed using Sequencher computer software (Gene Codes Corporation, Ann Arbor, MI).

Statistical Methods. HWE tests for all SNPs and LD tests for all pairs of SNPs were performed using the method of exact tests as implemented in the Genetic Data Analysis (GDA) computer program (15). The empirical *P* were based on 10,000 replicate samples for Monte Carlo simulations.

Genotypic frequencies of each SNP were compared between cases and controls. The hypotheses of differences in genotypic frequencies (three genotypes) between cases and controls were tested using the FET. An unconditional logistic regression was used to test for association between genotypes and prostate cancer and to estimate the age-adjusted RR of risk genotypes (homozygous variant genotype *versus* homozygous wild-type genotype). ANOVA was used to test for differences in mean log PSA levels ($\log_{10}$ transformed) among men with different genotypes.

Family-based association tests were performed for a subset of SNPs in the 159 HPC families, using the FBAT software package (16). Briefly, FBAT calculates observed *S* statistics from the data, which is the linear combination of offspring genotypes and phenotypes. The distribution of the *S* statistics is generated by treating the offspring genotype data as random and conditioning the phenotypes and parental genotypes. A *Z* statistic and its corresponding *P* or an empirical *P* is calculated. The test for association is valid if the empirical variance is used to account for the correlation between transmissions in families when linkage is present.

All of the hypothesis tests were limited to Caucasians only, to decrease the impact of heterogeneity and potential population stratification.

Results

Eighteen *hOGG1* SNPs described in the Celera SNP database were selected for initial screening. Of these, two were not observed at all, and six were infrequently seen (the frequency of the less frequent allele, <0.05) in our first 96 samples and, thus, were not further genotyped in the rest of the samples. The remaining 10 SNPs were genotyped in the total 245 sporadic cases and 222 unaffected controls. All of the 10 SNPs were in HWE ($P < 0.05$), and all of the pair-wise SNPs were in strong LD ($P < 0.00001$) both in sporadic cases and in unaffected controls. When the genotype distributions of the 10 SNPs were compared between sporadic cases and controls (Table 2), three had differences in the genotype distributions (Ser326Cys, FET $P = 0.055$; 7143A/G, FET $P = 0.059$; 11657A/G, FET $P = 0.028$), although only the 11657A/G reached statistical significance.

These three SNPs were then further genotyped in 159 HPC probands (Table 2). The genotype distributions of Ser326Cys and 7143A/G in the HPC probands were similar to those in the controls (FET, $P = 0.34$ and 0.11, respectively). The distribution of 11657A/G in the HPC probands, however, was significantly different from that in the controls (FET, $P = 0.03$). Exploring the data, we found a higher frequency of *CC* homozygotes for the Ser326Cys and an especially higher frequency of *GG* homozygotes for the 11657A/G and 7143A/G in cases compared with controls. For example, there were 17 *GG* homozygotes at 11657A/G among 357 sporadic or HPC probands and only one *GG* homozygote in the 187 controls, although the subject had an elevated PSA level (3.9 ng/ml). Compared with men with the *AA* genotype at 11657A/G, men with the *GG* genotype were at increased risk for prostate cancer, even after adjustment for age. The point estimate of the RR was 9.80 (95% CI, 1.25–76.92) for sporadic prostate cancer, 13.89 (95% CI, 1.57–125) for hereditary prostate cancer, and 9.80 (95% CI, 1.30–76.92) for either type of prostate cancer (Table 3). Similar results were observed for the SNP 7143A/G. For the SNP Ser326Cys, men with the *CC* genotype (Ser326) had an increased risk of prostate cancer, especially sporadic prostate cancer, compared with homozygous *GG* men (Cys326). The estimated RR was 3.23 (95% CI, 1.19–8.73), 2.07 (95% CI, 0.65–6.62), and 2.72 (95% CI, 1.17–6.32), for sporadic, hereditary, and either type of prostate cancer, respectively.

Because cases and controls may come from different genetic backgrounds, and any observed genotypic difference may reflect variation in genetic characteristics, rather than a difference directly related to the disease phenotype (*i.e.*, a population stratification effect), we performed a family-based association test to further examine the association between the sequence variants and prostate cancer risk, independent of potential population stratification. The SNPs

Table 2 *Genotype frequencies of sequence variants of hOGG1 in cases and controls (Caucasians only)*

SNPs	Genotype	No. of subjects (%)			*Ps* (*vs.* control)[a]	
		Controls	Sporadic	HPC	Sporadic	HPC
3402G/A	*AA*	79 (43)	73 (40)			
	AG	83 (45)	81 (44)			
	GG	23 (12)	29 (16)		N.S.[b]	
3574G/A	*AA*	104 (60)	128 (67)			
	AG	58 (34)	54 (28)			
	GG	11 (6)	10 (5)		N.S.	
6170G/C	*CC*	101 (58)	130 (63)			
	CG	60 (34)	69 (33)			
	GG	13 (7)	8 (4)		N.S.	
6803C/G (Ser326Cys)	*CC*	96 (55)	122 (61)	60 (61)		
	CG	63 (36)	71 (36)	35 (35)		
	GG	15 (9)	6 (3)	4 (4)	0.055	0.34
7143A/G	*AA*	130 (71)	153 (68)	83 (64)		
	AG	52 (28)	59 (26)	41 (32)		
	GG	2 (1)	12 (5)	6 (5)	0.059	0.11
9110A/G	*GG*	110 (60)	138 (66)			
	GA	62 (34)	66 (31)			
	AA	12 (7)	6 (3)		N.S.	
10629C/G	*CC*	53 (30)	54 (28)			
	CG	73 (41)	84 (44)			
	GG	53 (30)	51 (27)		N.S.	
10660A/T	*TT*	111 (61)	140 (65)			
	TA	59 (32)	69 (32)			
	AA	12 (7)	8 (4)		N.S.	
11657A/G	*AA*	139 (74)	158 (70)	88 (67)		
	AG	47 (25)	56 (25)	38 (29)		
	GG	1 (1)	11 (5)	6 (5)	0.028	0.03
11826A/T	*AA*	110 (60)	138 (66)			
	AT	60 (33)	64 (31)			
	TT	12 (7)	7 (3)		N.S.	

[a] FET.
[b] N.S., not significant.

11657A/G and 7143A/G were genotyped in all of the available family members of HPC families because the distribution of these two SNPs in the probands were significantly different from those in the controls. Parents who are heterozygous *A/G* for 11657A/G preferably transmit the *G* allele to affected sons (observed and expected *S* of 91 and 81, respectively; $Z = 2.28$, $P = 0.02$). A similar trend was observed for 7143A/G, although it was not statistically significant ($Z = 1.36$, $P = 0.17$). These results suggest that the observed differences of genotype distributions at 11657A/G between cases and controls are not solely attributable to the impact of population stratification.

We also tested the hypotheses that the sequence variants in *hOGG1* are associated with clinical characteristics of prostate cancer or preoperative PSA levels. When we compared the distributions of the 10 SNPs in sporadic cases with high (≥7) or low (≤6) Gleason scores and with a confined or nonlocalized tumor, no statistically significant difference in the genotypic frequencies of these SNPs was found between any of these groups (data not shown). We also compared the mean $\log_{10}$ PSA levels by the genotypes in these 10 SNPs among cases (preoperative) and controls, respectively. No significant difference was found in any of the groups.

Discussion

Although multiple functional studies have clearly demonstrated that *hOGG1* plays a critical role in repairing the major lesion 8-OH-G, limited data are available on the association between the sequence variants of the *hOGG1* and cancers. In this study, we provided new data to address this issue in prostate cancer. Our study is the first one to evaluate the sequence variants of *hOGG1* and prostate cancer risk using a comprehensive approach. Not only did we evaluate the previously reported missense change (Ser326Cys), but we also screened an additional 17 sequence variants spanning the entire gene, and we evaluated a total of 10 SNPs in the 245 sporadic cases and 222 unaffected controls. Furthermore, based on the results of sporadic cases and controls, we genotyped three SNPs with evidence for association in an additional 159 HPC probands. Most importantly, we applied family-based association tests to evaluate two of the three SNPs, to eliminate any potential impact of population stratification. We found that men with homozygous *G* at either 11657A/G or 7143A/G or with homozygous *C* (Ser326) at Ser326Cys, were at increased risk for prostate cancer, especially for sporadic prostate cancer. The finding of significant differences in the genotype distribution of 11657A/G between cases and controls was confirmed and significantly strengthened by the observation that heterozygous parents preferably transmit the *G* allele to affected sons, from a family-based association test. Taking these results together, our study provides strong preliminary evidence that sequence variants of *hOGG1* are associated with prostate cancer risk.

Although the significantly increased frequency of men homozygous for *G* at 11657A/G and *C* (Ser326) at Ser326Cys in both sporadic and hereditary cases, compared with controls, may be potentially attributable to random genotype error and/or population stratification, these confounding factors are unlikely to be major problems in our study for the following reasons: (*a*) the genotyping error rate should be very low in our study. A rigorous quality control is implemented in our genotyping laboratory by including both case and control samples in the same 384-well plates, the incorporation of multiple Centre d'Etude du Polymorphisme Humain (CEPH) controls in each plate, the use of robots in each step, and allele determination by a computer program. If genotyping error exists after these steps, it should be random to cases and controls. Furthermore, almost complete matching of the genotypes at 11657A/G and 7143A/G (caused by almost complete LD between these two SNPs) suggests a high quality of genotyping; and (*b*) potential population stratification, which is an inherent problem of any case-control study, is unlikely to play a major role in our findings. Our family-based linkage disequilibrium test, which is not susceptible to this confounding factor, provided the same significant finding for the SNP 11657A/G.

However, caution should be used when interpreting and generalizing these findings. The study subjects were recruited primarily for genetic studies rather than for a rigorously designed epidemiological study, thus making it difficult to generalize the point estimates of the RR. Furthermore, the control subjects, who were recruited from a prostate cancer screening population, are subject to potential misclassification in that they may represent a higher

Table 3 *Estimated RR of hOGG1 SNPs for prostate cancer (Caucasians only)*

SNPs	Genotype	RR (95% CI),[a] FET *P*		
		Sporadic	Hereditary	Either type of prostate cancer
6803C/G (Ser326Cys)	*GG*	1	1	1
	CC	3.23 (1.19–8.73) FET $P = 0.02$	2.07 (0.65–6.62) FET $P = 0.21$	2.72 (1.17–6.32) FET $P = 0.02$
7143A/G	*AA*	1	1	1
	GG	5.12 (1.12–23.25) FET $P = 0.03$	8.19 (1.51–45.45) FET $P = 0.06$	5.21 (1.18–22.73) FET $P = 0.03$
11657A/G	*AA*	1	1	1
	GG	9.80 (1.25–76.92) FET $P = 0.008$	13.89 (1.57–125.00) FET $P = 0.02$	9.80 (1.30–76.92) FET $P = 0.009$

[a] Adjusted for age.

risk population because of self-selection. This potential bias, however, is unlikely to be significant in our study, because very few of the 182 personally interviewed controls reported a positive family history (defined as an affected father and/or brothers). In addition, all of the control subjects were found to have normal DRE and PSA results at the time of screening. Lastly, we cannot rule out the impact of random sampling variation as a potential reason for our significant findings, especially when considering the low frequency of *GG* homozygotes for 11657A/G. Although we observed a higher frequency of *GG* homozygotes for 11657A/G and *CC* for Ser325Cys in both sporadic cases and hereditary cases, they were both compared with a single control group. Although replication of these findings in independent studies can definitively address this issue, the similar results observed in our family-based association study alleviate this concern substantially.

Although our results on the SNP Ser326Cys are unexpected, they are still consistent with the results from functional and epidemiological studies. The exact repair function associated with this sequence variant is unknown. Whereas Kohno *et al.* (3) demonstrated that the *Cys326* allele was about 7-fold less capable of complementing a repair deficient strain than the *Ser326* allele in an *in vitro* functional complementation assay, Dherin *et al.* (4) did not observe significant differences in *OGG1* activity of OGG1-glutathione *S*-transferase (GST) fusion proteins *in vitro*. A recent study by Janssen *et al.* (17) found that DNA repair activity of *OGG1* in human lymphocytes is not dependent on the Ser326Cys variant. Furthermore, the repair activity associated with this sequence variant *in vivo* in normal human cells is not known.

Paralleling the results of the functional studies, the results from epidemiological studies on the association between this sequence variant and cancer risk are inconclusive. The sequence variant Ser326Cys in germ-line DNA has been studied in several lung, esophageal, and gastric cancer populations. Two observations can be summarized from these studies: (*a*) although inconclusive, there is evidence that this sequence variant may be associated with susceptibility to several different cancers. For lung cancer, Sugimura *et al.* (9) found that individuals homozygous for *G* (Cys326) were at significantly increased risk for lung squamous cell carcinoma and nonadenocarcinoma in a Japanese population. However, two other studies did not confirm this association (3, 7). In the German population, Wikman *et al.* (7) found a higher proportion of *CC* homozygotes (Ser326) among lung cancer patients (64.8%) than in the controls (57.1%). It is worth noting that the frequency of *CC* homozygotes (Ser326) in the cases and controls of Wikman's study (7) are similar to what we observed in our prostate cancer cases (61.3%) and controls (55.2%), respectively. For esophageal cancer, Xing *et al.* (10) found that *GG* (Cys326) homozygotes were at significantly increased risk for developing esophageal squamous cell carcinoma in a Chinese population; and (*b*) there are significant differences in the genotype distribution between different races and ethnicities. The proportion of homozygous *C* (Ser326) individuals is highest in Melanesians (74.5%), Hungarians (63.7%), and Germans (57.1%), lower in Australian Caucasians (39.9%), Japanese (27.7%), and Micronesians (25.8%), and lowest in Chinese (12%; Refs. 7, 9). With the limited sample in our study, we observed 13 *CC* homozygotes out of 15 controls among African Americans. Interestingly, the proportions of the homozygous *C* (Ser326) are coincident with the different prevalence rates of prostate cancer in these populations. Furthermore, from these limited data, it seems that Ser326 confers risk to cancer in Caucasian populations and Cys326 confers risk to cancer in Asian populations.

Another potential limitation of this study is the possibility that some unknown sequence variants were not evaluated. This is especially true among the HPC cases, because only three SNPs were evaluated. However, we genotyped 10 SNPs across this gene, and there is significant pair-wise LD in all of the SNPs. Therefore, it is reasonable to expect that any increased prostate cancer risk caused by unknown sequence variants across *hOGG1* would most likely be reflected indirectly by at least one of the genotyped SNPs. However, we recognize that sequencing the entire gene and promoter region offers a definitive approach to identifying all of the important sequence variants, independent of the limitations of genotyping.

The stronger association of *hOGG1* SNPs observed in sporadic cases, compared with hereditary cases, was an unexpected finding. Although we can hypothesize that these are low-penetrance sequence variants, this assumption alone is not a sufficient explanation, because we would expect to observe at least similar risk to sporadic and hereditary prostate cancer if the inherited sequence variants confer any risk. Therefore, we think that at least two additional factors may contribute to this finding. First, competing high-penetrance genes may account for a significant proportion of the hereditary prostate cancer cases, such that the contribution of a low-penetrance gene, such as *hOGG1*, is relatively small in hereditary prostate cancer. The second contributing factor may be the unequal statistical power provided by the relatively small sample size of hereditary prostate cancer probands included in our study ($n = 133$, Caucasians), compared with sporadic cases ($n = 229$).

In summary, our study provides evidence for an association between sequence variants of *hOGG1* and prostate cancer risk. Considering the importance of this gene and the complexities of the available results, we conclude that additional epidemiological and functional studies are warranted not only in prostate cancer but also in other cancers.

Acknowledgments

We thank all of the subjects who participated in this study.

References

1. Boiteux, S. and Radicella, J. P. The human *OGG1* gene: structure, functions, and its implication in the process of carcinogenesis. Arch. Biochem. Biophys., *377:* 1–8, 2000.
2. Greenblatt, M. S., Bennett, W. P., Hollstein, M., and Harris, C. C. Mutations in the *p53* tumor suppressor gene: clues to cancer etiology and molecular pathogenesis. Cancer Res., *54:* 4855–4878, 1994.
3. Kohno, T., Shinmura, K., Tosaka, M., Tani, M., Kim, S. R., Sugimura, H., Nohmi, T., Kasai, H., and Yokota, J. Genetic polymorphisms and alternative splicing of the *hOGG1* gene, that is involved in the repair of 8-hydroxyguanine in damaged DNA. Oncogene, *16:* 3219–3225, 1998.
4. Dherin, C., Radicella, J. P., Dizdaroglu, M., and Boiteux, S. Excision of oxidatively damaged DNA bases by the human α-hOgg1 protein and the polymorphic α hOgg1(Ser326Cys) protein which is frequently found in human populations. Nucleic Acids Res., *27:* 4001–4007, 1999.
5. Audebert, M., Radicella, J. P., and Dizdaroglu, M. Effect of single mutations in the *OGG1* gene found in human tumors on the substrate specificity of the Ogg1 protein. Nucleic Acids Res., *28:* 2672–2678, 2000.
6. Hardie, L. J., Briggs, J. A., Davidson, L. A., Allan, J. M., King, R. F., Williams, G. I., and Wild, C. P. The effect of *hOGG1* and glutathione peroxidase I genotypes and 3p chromosomal loss on 8-hydroxydeoxyguanosine levels in lung cancer. Carcinogenesis (Lond.), *21:* 167–172, 2000.
7. Wikman, H., Risch, A., Klimek, F., Schmezer, P., Spiegelhalder, B., Dienemann, H., Kayser, K., Schulz, V., Drings, P., and Bartsch, H. *hOGG1* polymorphism and loss of heterozygosity (LOH): significance for lung cancer susceptibility in a caucasian population. Int. J. Cancer, *88:* 932–937, 2000.
8. Hanaoka, T., Sugimurabf, H., Nagura, K., Ihara, M., Li, X. J., Hamada, G. S., Nishimoto, I., Kowalski, L. P., Yokota, J., and Tsugane, S. *hOGG1* exon7 polymorphism and gastric cancer in case-control studies of Japanese Brazilians and non-Japanese Brazilians. Cancer Lett., *170:* 53–61, 2001.
9. Sugimura, H., Kohno, T., Wakai, K., Nagura, K., Genka, K., Igarashi, H., Morris, B. J., Baba, S., Ohno, Y., Gao, C., Li, Z., Wang, J., Takezaki, T., Tajima, K., Varga, T., Sawaguchi, T., Lum, J. K., Martinson, J. J., Tsugane, S., Iwamasa, T., Shinmura, K., and Yokota, J. *hOGG1* Ser326Cys polymorphism and lung cancer susceptibility. Cancer Epidemiol. Biomark. Prev., *8:* 669–674, 1999.

10. Xing, D. Y., Tan, W., Song, N., and Lin, D. X. Ser326Cys polymorphism in *hOGG1* gene and risk of esophageal cancer in a Chinese population. Int. J. Cancer, *95:* 140–143, 2001.
11. Wood, R. D., Mitchell, M., Sgouros, J., and Lindahl, T. Human DNA repair genes. Science (Wash. DC), *291:* 1284–1289, 2001.
12. De Marzo, A. M., Marchi, V. L., Epstein, J. I., and Nelson, W. G. Proliferative inflammatory atrophy of the prostate: implications for prostatic carcinogenesis. Am. J. Pathol., *155:* 1985–1992, 1999.
13. Osterod, M., Hollenbach, S., Hengstler, J. G., Barnes, D. E., Lindahl, T., and Epe, B. Age-related and tissue-specific accumulation of oxidative DNA base damage in 7,8-dihydro-8-oxoguanine-DNA glycosylase (Ogg1) deficient mice. Carcinogenesis (Lond.), *22:* 1459–1463, 2001.
14. Xu, J., Zheng, S. L., Chang, B., Smith, J. R., Carpten, J. D., Stine, O. C., Isaacs, S. D., Wiley, K. E., Henning, L., Ewing, C., Bujnovszky, P., Bleeker, E. R., Walsh, P. C., Trent, J. M., Meyers, D. A., and Isaacs, W. B. Linkage of prostate cancer susceptibility loci to chromosome 1. Hum. Genet., *108:* 335–345, 2001.
15. Weir, B. S. Genetic Data Analysis II: Methods for Discrete Population Genetic Data. Sunderland, MA: Sinauer Association, Inc. Publishers, 1996.
16. Laird, N. M., Horvath, S., and Xu, X. Implementing a unified approach to family-based tests of association. Genet. Epidemiol., *19 (Suppl. 1)*: S36–S42, 2000.
17. Janssen, K., Schlink, K., Gotte, W., Hippler, B., Kaina, B., and Oesch, F. DNA repair activity of 8-oxoguanine DNA glycosylase 1 (*OGG1*) in human lymphocytes is not dependent on genetic polymorphism Ser326/Cys326. Mutat. Res., *486:* 207–216, 2001.

Germline mutations and sequence variants of the macrophage scavenger receptor 1 gene are associated with prostate cancer risk

Jianfeng Xu[1], S. Lilly Zheng[1], Akira Komiya[2], Josyf C. Mychaleckyj[1], Sarah D. Isaacs[2], Jennifer J. Hu[1], David Sterling[3], Ethan M. Lange[1], Gregory A. Hawkins[1], Aubrey Turner[1], Charles M. Ewing[2], Dennis A. Faith[2], Jill R. Johnson[2], Hiroyoshi Suzuki[2], Piroska Bujnovszky[2], Kathleen E. Wiley[2], Angelo M. DeMarzo[2], G. Steven Bova[2], Baoli Chang[1], M. Craig Hall[1], David L. McCullough[1], Alan W. Partin[2], Vahan S. Kassabian[4], John D. Carpten[5], Joan E. Bailey-Wilson[5], Jeffrey M. Trent[5], Jill Ohar[1], Eugene R. Bleecker[1], Patrick C. Walsh[2], William B. Isaacs[2] & Deborah A. Meyers[1]

Published online: 16 September 2002, doi:10.1038/ng994

Deletions on human chromosome 8p22–23 in prostate cancer cells[1] and linkage studies in families affected with hereditary prostate cancer (HPC)[2–4] have implicated this region in the development of prostate cancer. The macrophage scavenger receptor 1 gene (*MSR1*, also known as *SR-A*) is located at 8p22 and functions in several processes proposed to be relevant to prostate carcinogenesis[5–10]. Here we report the results of genetic analyses that indicate that mutations in *MSR1* may be associated with risk of prostate cancer. Among families affected with HPC, we identified six rare missense mutations and one nonsense mutation in *MSR1*. A family-based linkage and association test indicated that these mutations co-segregate with prostate cancer ($P = 0.0007$). In addition, among men of European descent, *MSR1* mutations were detected in 4.4% of individuals affected with non-HPC as compared with 0.8% of unaffected men ($P = 0.009$). Among African American men, these values were 12.5% and 1.8%, respectively ($P = 0.01$). These results show that *MSR1* may be important in susceptibility to prostate cancer in men of both African American and European descent.

To evaluate the role of *MSR1* in prostate cancer susceptibility, we carried out a comprehensive genetic analysis using a large number of subjects from multiple populations. We first screened for sequence variants of *MSR1* in germline DNA samples from one affected individual (proband) from each of 159 families affected with HPC. We identified eight nonsynonymous changes, including one nonsense mutation at codon 293 (Arg293X), and seven missense mutations or sequence variants (Pro36Ala, Ser41Tyr, Val113Ala, Asp174Tyr, Pro275Ala, Gly369Ser and His441Arg). None of these sequence variants are listed in either the National Center for Biotechnology Information (NCBI) or Celera single nucleotide polymorphism (SNP) databases.

To investigate whether these mutations co-segregated with prostate cancer, we directly analyzed the sequences of all available DNA samples from all members of the original 159 families affected with HPC, as well as from each member of an additional 31 families with HPC that were identified subsequently (a total of 1,663, including 764 affected individuals). Whereas the missense variant Pro275Ala was found in 30 affected families, the other seven mutations were relatively rare and found in a total of only 13 families. We constructed pedigrees of the latter 13 families showing the mutation status and phenotypic information for each member (Fig. 1). The nonsense mutation Arg293X was observed in six different families (all of European descent), the missense change Asp174Tyr in four different families (all African American) and each remaining mutation in a single pedigree.

To test formally for co-segregation of the mutations and prostate cancer, we carried out parametric linkage analysis using the mutations as a combined biallelic marker. Moderate evidence for linkage as measured by a log likelihood ratio assuming heterogeneity (hlod score) of 1.73 was obtained ($P = 0.005$). Nonparametric analyses gave a similar result ($Z = 2.16, P = 0.02$). As an additional test for co-segregation, we implemented a family-based linkage and association analysis using the FBAT computer program[11], an approach with increased power to detect co-segregation when an association exists between a mutation and the disease. The S statistic that we calculated for the combined mutant alleles ($Z = 3.40$, $P = 0.0007$) was higher than expected, supporting a linkage between these mutations and prostate cancer. In contrast with the rare *MSR1* mutations, the common variant Pro275Ala did not show co-segregation with prostate cancer by analysis with FBAT ($Z = 0.11, P = 0.91$). Although these data provide statistical evidence for linkage between prostate cancer and the rare *MSR1* mutations, the overall evidence is modest; a clear pattern of co-segregation with prostate cancer was readily observable in some pedigrees but obviously lacking in others. In addition, in some pedigrees only a few of the affected members (for example, a single branch of a pedigree) carried mutations in *MSR1*, suggesting that additional high risk factors are involved in these families.

[1]*Center for Human Genomics and the Departments of Public Health, Medicine, Pediatrics, Urology and Cancer Biology, Wake Forest University School of Medicine, Winston-Salem, North Carolina 27157, USA.* [2]*Brady Urological Institute and the Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins Medical Institutions, 600 N. Wolfe Street, Baltimore, Maryland 21287, USA.* [3]*Department of Medicine, St Louis University, Missouri, USA.* [4]*Georgia Urology, Atlanta, Georgia, USA.* [5]*National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland, USA. Correspondence should be addressed to W.B.I. (e-mail: wisaacs@jhmi.edu).*

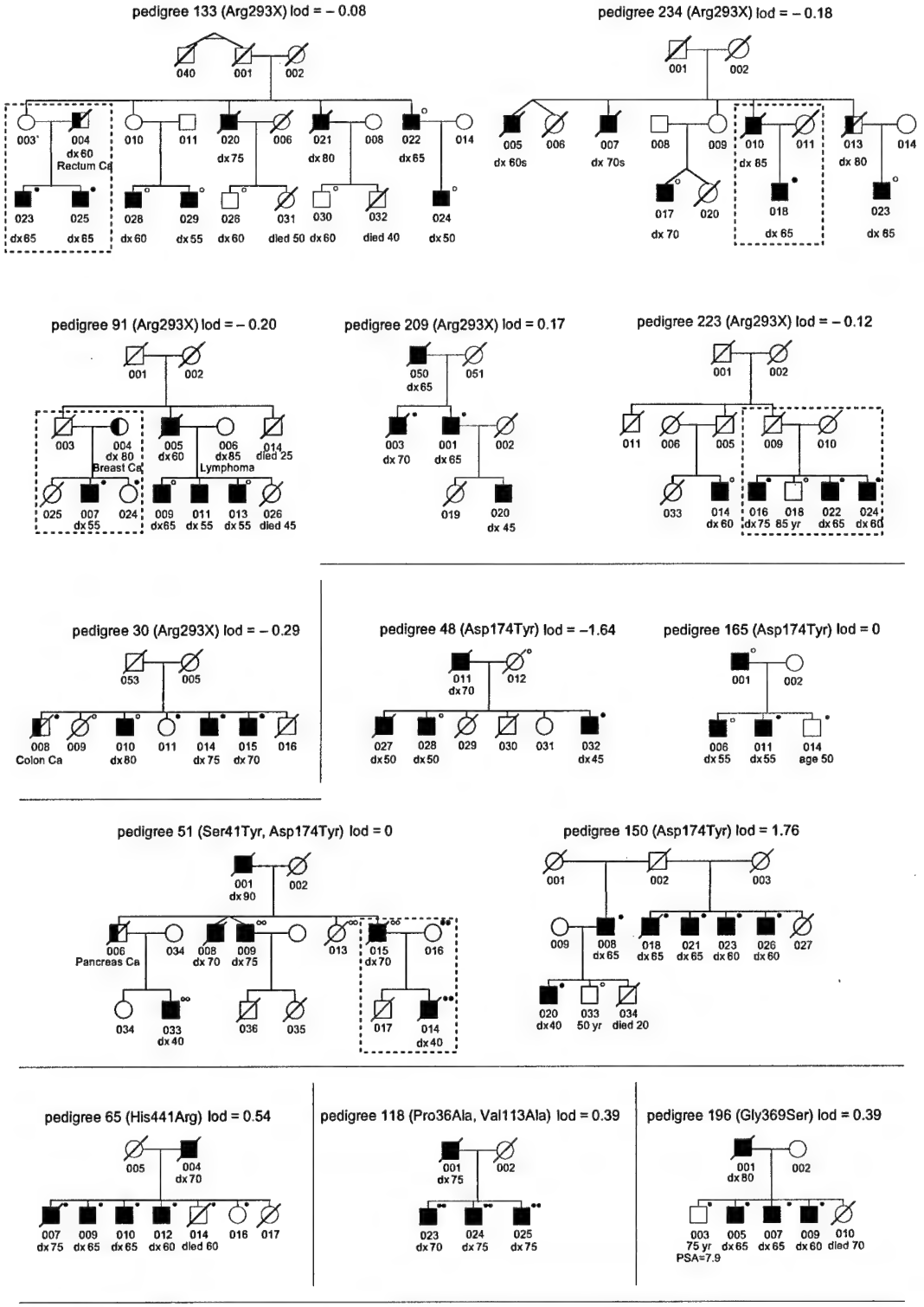


Fig. 1 Pedigrees representing the 13 families with *MSR1* mutations identified in this study (with minor changes in family structure to protect confidentiality). Fully filled boxes represent men affected with prostate cancer. Half-filled boxes and circles indicate men and women, respectivel y, with other types of cancer. Open boxes and circles represent unaffected men and women, respectively. Deceased individuals are indicated with a line bisecti ng the box or circle. Dashed boxes indicate five nuclear families from large kindreds affected with prostate cancer who segregated *MSR1* gene variants that were not found in the extended families. A superscripted circle indicates that a DNA sample from that individual was available and their genotype is known; a filled superscripted circle indicates carriers of the variant listed above the family pedigree, and an open superscripted circle indicates noncarriers. In families 5 1 and 118, two superscripted circles denote carriers of two variants. Although the pedigree structures have been altered to conceal the identity of families and stu dy participants, these alterations do not hinder interpretation of the segregation of *MSR1* alleles with prostate cancer. Ages were rounded to the nearest 5-year interval.

Table 1 • Rare *MSR1* germline mutations in individuals affected with prostate cancer and in unaffected controls

Race	Mutations	Non-HPC men	Unaffected men	Fisher exact test [a]
European descent		*n* = 317	*n* = 256	
	Arg293X	8	1	0.047
	Pro36Ala	0	0	–
	Ile54Val	1	0	1.00
	Val113Ala	2	1	1.00
	Gly369Ser	0	0	–
	His441Arg	3	0	0.26
African American		*n* = 48	*n* = 110	
	Ser41Tyr	2	1	0.22
	Asp174Tyr	6	2	0.01

[a]*P* values based on two-sided tests.

To examine whether the rare*MSR1* mutations contributed to our previous evidence for linkage of prostate cancer to 8p22–23 (ref. 2), we re-analyzed the linkage data using the same 24 markers to compare families with and without*MSR1* mutations from the original study. The 11 families with*MSR1* mutations had significantly higher lod scores (hlod = 1.40,*P* = 0.01) in this region than did the families without mutations (hlod = 0.05), suggesting that the former group contributes disproportionately to the overall linkage at 8p22–23. The proportion of families showing linkage to *D8S1135*, the closest microsatellite marker to*MSR1*, was significantly higher in the 11 families with*MSR1* mutations than in those without (45% versus 2%$\chi_1^2 = 4.28, P = 0.038$).

We evaluated further the association between these mutations and prostate cancer by screening a group of men with non-HPC (that is, affected men either without a family history of prostate cancer or with only one affected first-degree relative) and unaffected men. The nonsense mutation Arg293X was again found only in subjects of European descent and was observed significantly more frequently in individuals with non-HPC (8 individuals, 2.52%) than in unaffected men (1 individual, 0.39%,*P* = 0.047; Table 1). The one unaffected man carrying the Arg293X mutation was 65 years old and had a serum prostate-specific antigen (PSA) concentration of 2.1 ng ml^{-1}. Notably, of the 91 individuals in this group with seminal vesicle invasion or lymph node metastasis at the time of diagnosis, 6 (6.59%) carried Arg293X mutations, compared with 2 of the remaining 226 individuals (0.88%) who did not have evidence of metastatic or locally invasive disease (Fisher's exact test, *P* = 0.008).

Of the other rare *MSR1* mutations observed in families of European descent affected with HPC, only the His441Arg and Val113Ala mutations were observed in individuals with non-HPC (3 and 2 individuals, respectively). The His441Arg mutation was not observed in unaffected men, and the Val113Ala mutation was found in one unaffected man. Another mutation, Ile54Val, was observed in one individual with non-HPC. When these mutations were analyzed together with the Arg293X mutation, their combined frequency in individuals with non-HPC (4.42%) was significantly higher than their frequency in unaffected men (0.78%,*P* = 0.009; Table 1).

The missense mutation Asp174Tyr, found in four families affected with HPC, was observed only in African American subjects, where it occurred more often in individuals with non-HPC (6 individuals, 12.50%) than in unaffected men (2 individuals, 1.82%; *P* = 0.01; Table 1). The two unaffected men carrying the Asp174Tyr mutation (aged 56 and 60) both had a positive family history of prostate cancer, although their PSA concentrations were normal.

To determine the frequency and the impact of the two recurrent mutations, Arg293X and Asp174Tyr, in the general population, we screened an additional 518 men who had been selected for exposure to asbestos regardless of their prostate cancer status[12]. In this group of men, the nonsense mutation Arg293X was observed in 7 of 469 men of European descent (1.5%), a frequency that is intermediate between those observed for individuals with non-HPC and unaffected men. Notably, two carriers of the Arg293X mutation were among the 28 men in this group diagnosed with prostate cancer (7.1%). Two other elderly men

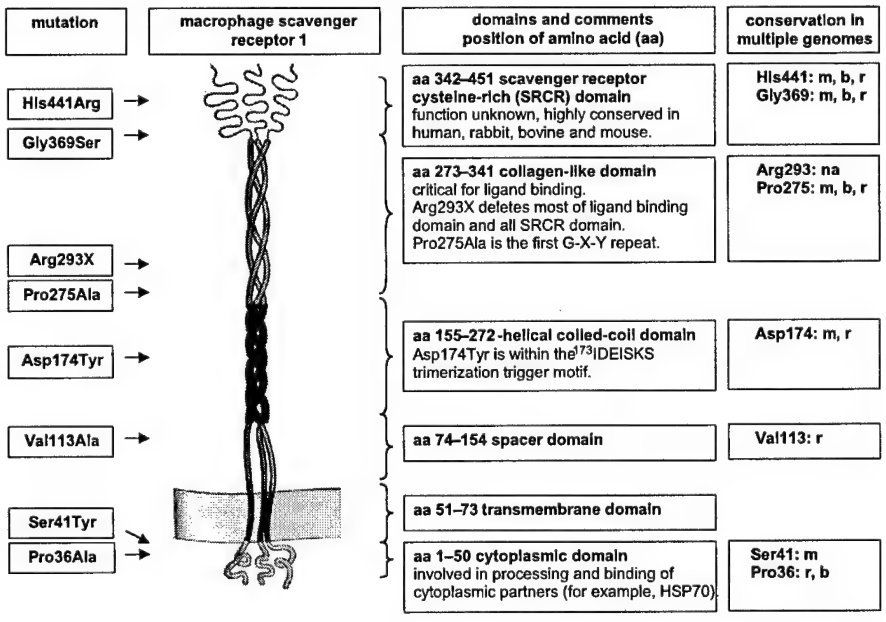


Fig. 2 Macrophage scavenger receptor 1. The locations of the mutations identified in this study are shown on the left. The functional domains and corresponding positions of the amino acids, and the results from protein alignment of multiple genomes, are shown on the right. MSR1 has been sequenced in four species: *Homo sapiens*, *Mus musculus* (mouse, m), *Oryctolagus cuniculus* (rabbit, r) and *Bos taurus* (cow, b). na, not applicable.

(aged 72 and 76) carried the mutation and had increased PSA concentrations (11.8 and 4.2 ng ml^{-1}, respectively). No biopsy data were available for these two men. Three other carriers were not diagnosed with prostate cancer and had normal PSA concentrations (1.5 ng ml^{-1} at age 58, 0.8 ng ml^{-1} at age 62, and 1.8 ng ml^{-1} at age 74). The missense mutation Asp174Tyr was observed in 2 of 49 unaffected African American men (4.1%); one had a PSA concentration of 3.4 ng ml^{-1} at age 64, and the other a PSA concentration of 0.4 ng ml^{-1} at age 43.

These mutation frequencies in men with and without cancer should be interpreted with caution, as the observed association is subject to potential population stratification. We think, however, that stratification would be minimal in this population because (i) the comparisons were carried out separately for subjects of European and African American descent; (ii) a sample of 24 consecutive SNPs on chromosomes 1, 8, 11, 12 and X that were genotyped in this population showed no evidence of population stratification (data not shown); and (iii) a family-based linkage and association test, which is insensitive to population stratification, statistically supported a role for *MSR1* in prostate cancer susceptibility.

The MSR1 protein is homotrimeric and has six predicted protein domains: the amino-terminal cytoplasmic domain, transmembrane domain, spacer domain, α-helical coiled-coil domain, collagen-like domain, and the scavenger receptor cysteine-rich carboxy-terminal domain (ref. 13; Fig. 2). This macrophage-specific receptor can bind many different polyanionic ligands, ranging from Gram-negative and Gram-positive bacteria, to oxidized low-density lipoprotein, to silica (reviewed in ref. 5). The truncating mutation Arg293X results in deletion of most of the collagen-like domain, including the ligand-binding region and the cysteine-rich domain[5,14]. Synthetic mutant MSR1 proteins that are similar to the predicted product of the Arg293X mutation have a dominant-negative phenotype when expressed *in vitro*[15,16]. Regarding the missense mutation Asp174Tyr (observed in African American individuals), mutagenesis studies have identified a crucial heptapeptide sequence, 173IDEISKS, in the α-helical coiled-coil domain of MSR1 that acts as the functional 'trigger' for proper polymerization of the three MSR1 polypeptide chains[17].

We have shown by immunochemical analysis that macrophages present in both benign and cancerous prostate tissues routinely express MSR1 (C.M.E., A.M.D. and W.B.I., unpublished observations). Inflammation and features such as proliferative regeneration of prostate epithelium in the presence of increased oxidative stress that are associated with this expression probably have key roles in the development of prostate cancer[6]. MSR1, through its induction by oxidative stress[18] and its ability to bind oxidized low-density lipoprotein, may modify amounts of reactive oxygen intermediates in this context. The finding that *MSR1* knockout mice have a reduced capacity to eradicate certain pathogens effectively may also be relevant[19,20], because an infectious etiology of prostate cancer has been proposed[21].

In summary, we have presented genetic evidence showing that *MSR1* may have an important role in susceptibility to prostate cancer. Given the modest amount of evidence, however, follow-up studies are necessary to verify the associations observed in this study.

Methods

Subjects. The subjects studied were from four different populations. The first group comprised 159 families affected with HPC who were recruited either at the Brady Urology Institute at Johns Hopkins Hospital (Baltimore, Maryland) through referrals or review of medical records for individuals attending the hospital for treatment of prostate cancer, or as respondents to various lay publications describing our studies[2]. Each family had at least three first-degree relatives affected with prostate cancer. Diagnosis was verified by medical records. The mean age at prostate cancer diagnosis for these probands was 61 years. The men included 133 (84%) of European descent and 14 (8.8%) African Americans. We subsequently added another 31 families with HPC, which were identified in the same way.

The second group comprised unrelated men affected with non-HPC and unaffected men. Men with non-HPC (335) were recruited from individuals attending Johns Hopkins Hospital for treatment of prostate cancer and comprised 317 men of European descent and 18 African Americans. Diagnosis was confirmed by pathology reports for each subject. The mean age at prostate cancer diagnosis for this group was 58.6 years. Unaffected individuals (346) were recruited from men participating in screening programs for prostate cancer who had normal digital rectal examination results and normal PSA concentrations (< 4 ng ml^{-1}) and comprised 256 men of European descent and 90 African Americans. The mean age at examination was 58 years, and 5.8% of men in this population had a father or brother affected with prostate cancer.

The third group was a small African American case-control population from Wake Forest University School of Medicine, added to this study to enlarge the sample size of African Americans. Among these men, 30 individuals were affected with prostate cancer, and 20 individuals were unaffected men who participated in screening programs, were at least 50 years of age and had normal digital rectal examination results and PSA concentrations.

The fourth group was a subset (n = 518) of a large population study of workers exposed to asbestos who were recruited to study the impact of genetic and environmental factors on the development of asbestos-induced lung diseases[12]. The racial distribution of these men was 91% of European and 9% of African American descent. The mean age of the study subjects at examination was 63.6 years. Prostate cancer diagnosis was reported in 6.0% (n = 31) of men; this rate was similar in men of European descent (5.9%, n = 28) and African Americans (6.1%, n = 3). Serum concentrations of PSA and prostate cancer diagnoses were obtained subsequently. Participants worked as painters, pipe fitters, plumbers, operators and electricians. A physical examination was carried out on all participants. The Institutional Review Boards of Johns Hopkins University, St Louis University and Wake Forest University approved each of the study protocols. Informed consent was obtained from all subjects who participated in the study.

Sequence analysis and SNP genotype analysis. We directly determined the sequences of the PCR products of all 11 *MSR1* exons, exon–intron junctions, promoter regions and 5′ and 3′ untranslated regions (UTRs; ref. 22). The primers used for PCR are available from W.B.I. on request. All PCR reactions were done in a volume of 10 μl containing 30 ng of genomic DNA, each primer at 0.2 μM, each dNTP at 0.2 mM, 1.5 mM $MgCl_2$, 20 mM Tris–HCl, 50 mM KCl and 0.5 U of *Taq* polymerase (Life Technologies). PCR cycling conditions were as follows: 94 °C for 4 min; 30 cycles of 94 °C for 30 s, the specified annealing temperature for 30 s, and 72 °C for 30 s; and a final extension of 72 °C for 6 min. All PCR products were purified using the QuickStep PCR purification kit (Edge BioSystems) to remove dNTPs and excess primers. We carried out all reactions using dye-terminator chemistry (BigDye, ABI) and used 63 ± 5% ethanol for precipitation. We loaded samples onto an ABI 3700 DNA Analyzer after adding 8 μl of formamide. SNPs were identified using Sequencher software version 4.0.5 (Gene Codes Corporation).

Computational analysis. We assembled the complete human mRNA sequence corresponding to the type I and II isoforms of *MSR1* by optimal pair-wise alignment of mRNA subsequences using the GCG Bestfit program (Accelrys). Only the coding sequence of type III was available in GenBank. Exon–intron boundaries in the NCBI sequence of human genome chromosome 8 were delineated by Smith–Waterman alignment of assembled type I, II, and III mRNA sequences to the human genome sequence using the Swat program (P. Green, unpublished data). We analyzed the secondary structure of the proteins by GCG programs, and predicted transmembrane regions with HMMTOP 2.0 (ref. 23) and TMHMM 2.0 (ref. 24).

Accession numbers. Nucleotide: D13263, human *MSR1* promoter and exon 1; D90187, type I mRNA coding sequence; D13264, type I 3′ UTR sequence; D90188, type II mRNA coding sequence; D13265, type II 3′ UTR sequence; AF037351, type III coding sequence. Peptide: BAA14298, MSR1 type I protein sequence; BAA14299, MSR1 type II protein sequence; AAC09251, type III protein sequence. Genomic: NT_015280.5, human genome chromosome 8 sequence contig.

Statistical analysis. We used the GDA computer program[25] to do Hardy–Weinberg Equilibrium (HWE) tests for all SNPs, and linkage disequilibrium (LD) tests for all pairs of SNPs. Linkage analyses used both parametric and nonparametric methods, implemented by the computer program GENEHUNTER[26]. For the parametric analysis, an autosomal dominant model that had been used previously[27] was assumed. We assessed linkage in the presence of heterogeneity using Smith's admixture test for heterogeneity[28]. We used a maximum likelihood approach to estimate the proportion of linked families (α) by maximizing the admixed lod score (hlod). We used a likelihood ratio test to test for different proportions of linked families (α values) between two groups of families, and calculated χ^2 according to $\chi^2 = 4.6 \times (\text{hlod}_1 + \text{hlod}_2 - \text{hlod}_{\text{total}})$ with 1 d.f., where hlod_1, hlod_2 and $\text{hlod}_{\text{total}}$ are the hlod scores for the two subsets of families and the whole sample, respectively. We used the statistic '*Z*-all' in the program for the non-mode-of-inheritance analysis[29].

To test for co-segregation between the rare mutations and prostate cancer, we constructed a biallelic marker by coding all seven different rare mutations into one mutation. We used FBAT software to do family-based linkage and association tests[11]. FBAT uses data from nuclear families and sibling relationships to determine an *S* statistic, which is the linear combination of offspring genotypes and phenotypes. The distribution of the *S* statistic is generated by treating the offspring genotype data as random and conditioning on the phenotypes and parental genotypes. A *Z* statistic and its corresponding *P* value are calculated. The hypotheses of differences in allele frequencies between cases and controls were tested on the basis of the χ^2 of Amitage trend tests[30], with adjustment for age.

Acknowledgments

We thank all subjects for their participation; F. Torti for his support in recruiting study subjects at Wake Forest University School of Medicine; W. Catalona for measuring PSA concentrations in the asbestos study population; R. Gurganus, L. Mangold and D. Lamm for help in prostate cancer screenings; and J.L. Hicks for help with immunohistochemistry. This work was partially supported by US Public Health Service Prostate Cancer Specialized Programs of Research Excellence grant, grants from the US Department of Defense to W.B.I. and J.X., Association for the Cure of Cancer of the Prostate, the Fund for Research and Progress in Urology, Johns Hopkins University, Wake Forest Comprehensive Cancer Center and the William Thomas Gerrard, Mario Anthony Duhon, Jennifer and John Chalsty Professorship in Urology.

Competing interests statement

The authors declare that they have no competing financial interests.

Received 22 May; accepted 19 August 2002.

1. Latil, A. & Lidereau, R. Genetic aspects of prostate cancer.*Virchows Arch.* **432**, 389–406 (1998).
2. Xu, J. *et al.* Linkage and association studies of prostate cancer susceptibility: evidence for linkage at 8p22–23.*Am. J. Hum. Genet.* **69**, 341–350 (2001).
3. Gibbs, M. *et al.* A genomic scan of families with prostate cancer identifies multiple regions of interest.*Am. J. Hum. Genet.* **67**, 100–109 (2000).
4. Goddard, K.A., Witte, J.S., Suarez, B.K., Catalona, W.J. & Olson, J.M. Model-free linkage analysis with covariates confirms linkage of prostate cancer to chromosomes 1 and 4.*Am. J. Hum. Genet.* **68**, 1197–1206 (2001).
5. Platt, N. & Gordon, S. Is the class A macrophage scavenger receptor (SR-A) multifunctional?—The mouse's tale.*J. Clin. Invest.* **108**, 649–654 (2001).
6. DeMarzo, A.M., Marchi, V.L., Epstein, J.I. & Nelson, W.G. Proliferative inflammatory atrophy of the prostate: implications for prostatic carcinogenesis. *Am. J. Pathol.* **155**, 1985–1992 (1999).
7. Nelson, W.G., DeMarzo, A.M. & DeWeese, T.L. The molecular pathogenesis of prostate cancer: implications for prostate cancer prevention.*Urology* **57**, 39–45 (2001).
8. Shimura, S. *et al.* Reduced infiltration of tumor-associated macrophages in human prostate cancer: association with cancer progression.*Cancer Res.* **60**, 5857–5861 (2000).
9. Lissbrant, I.F. *et al.* Tumor-associated macrophages in human prostate cancer: relation to clinicopathological variables and survival.*Int. J. Oncol.* **17**, 445–451 (2000).
10. Bova, G.S. *et al.* Homozygous deletion and frequent allelic loss of chromosome 8p22 loci in human prostate cancer.*Cancer Res.* **53**, 3869–3873 (1993).
11. Laird, N.M., Horvath, S. & Xu, X. Implementing a unified approach to family-based tests of association.*Genet. Epidemiol.* **19** (Suppl. 1), S36–S42 (2000).
12. Xu, J. *et al.* Association studies of serum prostate-specific antigen levels and the genetic polymorphisms at the androgen receptor and prostate-specific antigen genes. *Cancer Epidemiol. Biomarkers Prev.***11**, 664–669 (2002).
13. Kodama, T. *et al.* Type I macrophage scavenger receptor contains alpha helical and collagen-like coiled domains.*Nature* **343**, 531–535 (1990).
14. Wada, Y. *et al.* Structure and function of macrophage scavenger receptors.*Ann. NY Acad. Sci.* **748**, 226–238 (1995).
15. Acton, S. *et al.* The collagenous domains of macrophage scavenger receptors and complement component C1q mediate their similar, but not identical, binding specificities for polyanionic ligands.*J. Biol. Chem.* **268**, 3530–3537 (1993).
16. Dejager, S., Mietus-Snyder, M., Friera, A. & Pitas, R.E. Dominant negative mutations of the scavenger receptor. Native receptor inactivation by expression of truncated variants.*J. Clin. Invest.* **92**, 894–902 (1993).
17. Frank, S., Lustig, A., Schulthess, T., Engel, J. & Kammerer, R.A. A distinct seven-residue trigger sequence is indispensable for proper coiled-coil formation of the human macrophage scavenger receptor oligomerization domain.*J. Biol. Chem.* **275**, 11672–11677 (2000).
18. Mietus-Snyder, M., Glass, C.K. & Pitas, R.E. Transcriptional activation of scavenger receptor expression in human smooth muscle cells requires AP-1/c-Jun and C/EBPbeta: both AP-1 binding and JNK activation are induced by phorbol esters and oxidative stress.*Arterioscler. Thromb. Vasc. Biol.***18**, 1440–1449 (1998).
19. Suzuki, H. *et al.* A role for macrophage scavenger receptors in atherosclerosis and susceptibility to infection.*Nature* **386**, 292–296 (1997).
20. Thomas, C.A. *et al.* Protection from lethal gram-positive infection by macrophage scavenger receptor-dependent phagocytosis.*J. Exp. Med.* **191**, 147–156 (2000).
21. Strickler, H.D. & Goedert, J.J. Sexual behavior and evidence for an infectious cause of prostate cancer.*Epidemiol. Rev.* **23**, 144–151 (2001).
22. Emi, M. *et al.* Structure, organization, and chromosomal mapping of the human macrophage scavenger receptor gene.*J. Biol. Chem.* **268**, 2120–2125 (1993).
23. Tusnady, G.E. & Simon, I. The HMMTOP transmembrane topology prediction server.*Bioinformatics* **17**, 849–850 (2001).
24. Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E.L. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes.*J. Mol. Biol.* **305**, 567–580 (2001).
25. Weir, B.S. *Genetic Data Analysis II: Methods for Discrete Population Genetic Data* (Sinauer, Sunderland, Massachusetts, 1996).
26. Kruglyak, L., Daly, M.J., Reeve-Daly, M.P. & Lander, E.S. Parametric and nonparametric linkage analysis: a unified multipoint approach.*Am. J. Hum. Genet.* **58**, 1347–1363 (1996).
27. Smith, J. R. *et al.* Major susceptibility locus for prostate cancer on chromosome 1 suggested by a genome-wide search.*Science* **274**, 1371–1374 (1996).
28. Ott, J. *Analysis of Human Genetic Linkage* 3rd edn (Johns Hopkins Press, Baltimore, Maryland, 1998).
29. Whittemore, A.S. & Halpern, J. Probability of gene identity by descent: computation and applications.*Biometrics* **50**, 109–117 (1994).
30. Sasieni, P.D. From genotypes to genes: doubling the sample size.*Biometrics* **53**, 1253–1261 (1997).

Am. J. Hum. Genet. 72:208–212, 2003

Report

Common Sequence Variants of the Macrophage Scavenger Receptor 1 Gene Are Associated with Prostate Cancer Risk

Jianfeng Xu,[1] S. Lilly Zheng,[1] Akira Komiya,[2] Josyf C. Mychaleckyj,[1] Sarah D. Isaacs,[2] Baoli Chang,[1] Aubrey R. Turner,[1] Charles M. Ewing,[2] Kathleen E. Wiley,[2] Gregory A. Hawkins,[1] Eugene R. Bleecker,[1] Patrick C. Walsh,[2] Deborah A. Meyers,[1] and William B. Isaacs[2]

[1]Center for Human Genomics, Wake Forest University School of Medicine, Winston-Salem, and [2]Brady Urological Institute, Johns Hopkins Medical Institution, Baltimore

Rare germline mutations of macrophage scavenger receptor 1 (*MSR1*) gene were reported to be associated with prostate cancer risk in families with hereditary prostate cancer (HPC) and in patients with non-HPC (Xu et al. 2002). To further evaluate the role of *MSR1* in prostate cancer susceptibility, at Johns Hopkins Hospital, we studied five common variants of *MSR1* in 301 patients with non-HPC who underwent prostate cancer treatment and in 250 control subjects who participated in prostate cancer–screening programs and had normal digital rectal examination and PSA levels (<4 ng/ml). Significantly different allele frequencies between case subjects and control subjects were observed for each of the five variants (*P* value range .01–.04). Haplotype analyses provided consistent findings, with a significant difference in the haplotype frequencies from a global score test ($P = .01$). Because the haplotype that is associated with the increased risk for prostate cancer did not harbor any of the known rare mutations, it appears that the observed association of common variants and prostate cancer risk are independent of the effect of the known rare mutations. These results consistently suggest that *MSR1* may play an important role in prostate carcinogenesis.

The macrophage scavenger receptor 1 (*MSR1* [MIM 153622]) gene was recently identified as a candidate gene for prostate cancer susceptibility, from a systematic search for prostate cancer (MIM 176807) genes at 8p, using multiple approaches such as linkage, direct sequencing, and association studies (Xu et al. 2002). Analysis of the *MSR1* gene sequence in members of families with hereditary prostate cancer (HPC) identified six rare missense mutations (Pro36Ala, Ser41Tyr, Val113Ala, Asp174Tyr, Gly369Ser, and His441Arg) and one nonsense mutation (Arg293X). A family-based linkage and association test provided statistical evidence that these mutations cosegregate with prostate cancer ($P = .0007$). Further examination of these mutations in a collection of patients with non-HPC and unaffected men revealed that they were either not observed or were observed less frequently in men without prostate cancer. For white men, the rare *MSR1* mutations were detected in 4.4% of cases, compared to 0.8% in unaffected men ($P = .009$); for African American men, these values were 12.50% and 1.82%, respectively ($P = .01$). These results provide genetic evidence that *MSR1* may play an important role in prostate cancer susceptibility in both African American men and men of European descent.

The MSR1 protein, a Class A scavenger receptor, is a multidomain trimeric molecule composed of identical protein chains. It has two functional isoforms (Type I and Type II) and one nonfunctional isoform (Type III), generated by alternative splicing of a single 11-exon mRNA (Kodama et al. 1990; Emi et al. 1993). This macrophage-specific receptor is capable of binding a highly diverse array of polyanionic ligands, ranging from gram negative and positive bacteria and oxidized LDL

Received September 12, 2002; accepted for publication October 24, 2002; electronically published December 6, 2002.

Address for correspondence and reprints: Dr. Jianfeng Xu, Center for Human Genomics, Wake Forest University School of Medicine, Medical Center Boulevard, Winston-Salem, NC 27157, e-mail: jxu@wfubmc.edu; or Dr. Deborah A. Meyers, Center for Human Genomics, Wake Forest University School of Medicine, Medical Center Boulevard, Winston-Salem, NC 27157, e-mail: dmeyers@wfubmc.edu

0002-9297/2003/7201-0023$15.00

Table 1

Pairwise Linkage Disequilibrium (Correlation Coefficient) in Patients With Prostate Cancer and in Unaffected Control Subjects

	PRO3	INDEL1	IVS5-59	P275A	INDEL7
PRO3		.99[a]	.59[a]	.03	.04
INDEL1	.98[a]		.58[a]	.03	.04
IVS-59	.69[a]	.71[a]		.08	.08
P275A	.06	.05	.06		.86[a]
INDEL7	.03	.02	.07	.73[a]	

NOTE.—Estimates in the upper right are for control subjects and in the lower left are for case subjects.

[a] Indicates $P < .00001$, which was based on the permutation tests of exact test statistics using GDA.

to silica, and, correspondingly, has been linked to a wide variety of normal and pathological processes, including inflammation, innate and adaptive immunity, oxidative stress, and apoptosis (Platt and Gordon 2001). Although the exact role of *MSR1* in prostate carcinogenesis is unknown, some or all of these processes have been implicated in the development of prostate cancer (De Marzo et al. 1999; Nelson et al. 2001). Recent findings showing that the degree of macrophage infiltration is associated with prostate cancer prognosis strengthen the link between *MSR1* and prostate cancer (Lissbrant et al. 2000; Shimura et al. 2000).

Besides the seven rare *MSR1* mutations we reported elsewhere (Xu et al. 2002), we identified four additional common sequence variants (>10%) after sequencing the PCR products of all 11 exons, exon-intron junctions, promoter region, and 5′ and 3′ UTRs of *MSR1* in the germline DNA of probands from each of the 159 families with HPC (i.e., at least three first-degree relatives affected with prostate cancer). Together with the common missense change we identified elsewhere (Xu et al. 2002), there are five common sequence variants in our sequenced regions.

To evaluate the role of *MSR1* sequence variants in prostate cancer susceptibility as it relates to the general population, we focused this study on the common sequence variants and patients with prostate cancer who do not meet criteria for HPC. Because of the limited number of African Americans in our study, all the following analyses were restricted to men of European descent. Five common sequence variants were genotyped in 301 patients with prostate cancer and in 250 unaffected control subjects. The case subjects were recruited from the patients who underwent prostate cancer treatment at Johns Hopkins Hospital and did not meet the criteria for HPC ($N = 317$). The diagnosis of prostate cancer was confirmed by pathology reports, and the mean age at diagnosis was 59.3 years. Sixteen patients were excluded from the current study because of insufficient DNA. The control subjects were recruited from volunteers participating in prostate cancer–screening programs at Johns Hopkins Hospital who had a normal digital rectal examination and PSA (<4 ng/ml) ($N = 256$). The mean age at examination was 58.5 years. Six control subjects were excluded from the current study because of insufficient DNA. All subjects in this study gave full informed consent.

The five sequence variants genotyped in this study include an SNP in the promoter region (PRO3), a 15-bp insertion/deletion of "GAATGCTTTATTGTA" in intron 1 (INDEL1), an SNP in intron 5 (IVS5-59), a missense change in exon 6 (P275A), and a 3-bp insertion/deletion of "TTA" in intron 7 (INDEL7). The positions of these sequence variants are listed in table 2. Genotyping of the three SNPs (PRO3, IVS-59, and P275A) was performed using the MassARRAY system (SEQUENOM). Genotyping of the two insertion/deletions (INDEL1 and INDEL7) was performed using the 3700 DNA Analyzer (Applied Biosystems). The primer information and detailed PCR conditions for these sequence variants are available on the authors' Web site.

All five sequence variants were in Hardy-Weinberg equilibrium in control subjects, on the basis of 10,000 permutations of the Fisher probability test statistic (Weir 1996), as implemented in the software Genetic Data Anal-

Table 2

Frequencies of MSR1 Sequence Variants in Patients with Prostate Cancer and Unaffected Control Subjects

SNP, (POSITION),[a] AND GENOTYPE	NO. (%) OF SUBJECTS WITH GENOTYPE	
	Control Subjects	Case Subjects
PRO3 (-14,742):		
AA	214 (85.6)	235 (78.1)
AG	34 (13.6)	58 (19.3)
GG	2 (.8)	8 (2.6)
INDEL1[b] (-14,458):		
−/−	211 (85.4	233 (78.4)
-/+	33 (13.4)	58 (19.5)
+/+	3 (1.2)	6 (2.0)
IVS5-59 (22,788):		
CC	232 (92.8)	262 (87.3)
CA	18 (7.2)	34 (11.3)
AA	0 (.0)	4 (1.3)
P275A (22,850):		
CC	209 (83.6)	271 (90.3)
CG	38 (15.2)	28 (9.3)
GG	3 (1.2)	1 (.3)
INDEL7 (34,504):[c]		
−/−	204 (82.9)	264 (88.9)
−/+	41 (16.7)	33 (11.1)
+/+	1 (.4)	0 (.0)

[a] Positions (bp) are based on the initiation codon (ATG) from MSR1 genomic DNA (NT_015280).

[b] "+" and "−" denote with and without the 15-bp sequence "GAATGCTTTATTGTA," respectively.

[c] "+" and "−" denote with and without the 3-bp sequence "TTA," respectively.

Table 3

Allele Frequencies in Patients with Prostate Cancer and Unaffected Control Subjects

Allele	Allele Frequencies (%)		χ^2 Test for Allele (*P* Values)
	Control Subjects	Case Subjects	
PRO3 "G"	7.6	12.3	.01
INDEL1 "+"[a]	7.9	11.8	.04
IVS5-59 "A"	3.6	7.0	.02
P275A "C"	91.2	95.0	.01
INDEL7 "−"[b]	91.3	94.4	.04

[a] "+" denotes the presence of the 15-bp sequence "GAATGCTTTATTGTA."

[b] "−" denotes the absence of the 3-bp sequence "TTA."

ysis (GDA). The first three sequence variants (PRO3, INDEL1, and IVS5-59) were in strong linkage disequilibrium (LD), because the tests for pairwise LD among them were all highly significant (all $P < .00001$), again on the basis of 10,000 permutations of the exact test statistic (Weir 1996), as implemented in GDA. The last two variants also had strong LD between them ($P < .00001$). There was no LD between the blocks of the first three variants and the last two variants (all $P > .2$). The estimates of pairwise LD between all these sequence variants in case subjects and control subjects, as measured by correlation coefficients and implemented in SAS/Genetics, are presented in table 1.

The allele frequencies of the five sequence variants were all significantly different between case subjects and control subjects, on the basis of a χ^2 test for allele frequencies, with 1 df (table 2; table 3). Specifically, the frequency of allele "G" of PRO3 ($P = .01$), 15-bp insertion (+) of INDEL1 ($P = .04$), allele "A" of IVS5-59 ($P = .02$), allele "C" of P275A ($P = .01$), and deletion (−) of INDEL7 ($P = .04$), were higher in case subjects than in control subjects, respectively. To estimate the prostate cancer risk of these variants, we performed the tests by grouping three genotypes into two genotypes as shown in table 4. Except for the INDEL7, there were significant elevated risks for prostate cancer among the sequence variants, even when adjusted for age.

The haplotype analysis of these five sequence variants using the EM algorithm (Excoffier and Slatkin 1995) estimated that four major haplotypes account for >96% of all haplotypes (table 5). The haplotype frequencies were significantly different between case subjects and control subjects, with a *P* value of .011 from 10,000 simulations of global score tests, as implemented in haplo.score (Schaid et al. 2002). When specific haplotypes were examined, the haplotype "G (+) A C (−)" of these five variants (in the order of PRO3, INDEL1, IVS5-59, P275A, and INDEL7) had a significantly higher frequency in prostate cancer patients (6.6%) than in control subjects (2.6%), with a *P* value of .004 (on the basis of 10,000 simulations). It is worth noting that this haplotype did not harbor any of the rare mutations (Arg293X, His441Arg, Val113Ala, and Ile54Val) (Xu et al. 2002). For example, all eight occurrences of Arg293X and three occurrences of His441Arg resided on the haplotype "A (−) C C (−)." Therefore, it appears that the significant association between the common *MSR1* sequence variants and prostate cancer risk is independent of the impact of the known rare *MSR1* mutations.

Caution should be taken when interpreting these findings. Although the significant differences in allele and haplotype frequencies between patients with prostate cancer and unaffected control subjects could be due to the prostate cancer risk associated with these polymorphisms, it could also be due to other reasons, such as a type I error or population stratification. Regarding type I error, all the reported significance levels were nominal *P* values and were not adjusted for multiple comparisons. If we considered that at least 15 tests were performed in this report, and that the commonly suggested Bonferoni correction was used, none of the tests was significant at $P = .05$. However, the Bonferoni correction is not optimal in this case; not all of these tests were independent because of the LD between these polymorphisms and the dependence between allele and haplotype. Regarding population stratification, the results, as a case-control study, are always subject to this potential confounder: that is, the different genotype frequencies observed may partially reflect different genetic backgrounds in case subjects and control subjects. Although great attention was paid in the study design and analysis,

Table 4

Odds Ratio (OR) Estimates for Prostate Cancer, Adjusting for Age

Variant: Genotypes	OR (95% CI)
PRO3:	
AA	1.00
AG/GG	1.81 (1.15–2.85)
INDEL1:	
−/−	1.00
+/− or +/+	1.73 (1.10–2.72)
IVS-59:	
CC	1.00
CA/AA	1.93 (1.07–3.50)
P275A:	
CG/GG	1.00
CC	1.75 (1.05–2.94)
INDEL7:	
−/−	1.00
+/− or +/+	1.54 (.94–2.52)

NOTE.—OR estimates are adjusted for age.

Table 5

Haplotype Frequencies of MSR1 Sequence Variants in Patients with Prostate Cancer and Unaffected Control Subjects

Haplotype[a]	Frequency (95% CI)		*P* values (empirical)
	Control Subjects	Case Subjects	Haplotype Specific
A (−) C C (−)	83.6% (80.3–86.9)	81.8% (78.7–84.9)	.44
G (+) A C (−)	2.6% (1.2–4.0)	6.6% (4.6–8.6)	.004
A (−) C G (+)	6.5% (4.3–8.6)	4.0% (2.4–5.5)	.06
G (+) C C (−)	3.7% (2.1–5.4)	4.1% (2.5–5.7)	.77
Global			.011

[a] Haplotype of five SNPs (in the order of PRO3, INDEL1, IVS-59, P275A, and INDEL7)

including the restriction to subjects of European descent only, we cannot rule out the possibility of population stratification.

Each of these five sequence variants could have an important impact on MSR1 function. For example, the SNP in the promoter region and the 15-bp ins/del polymorphism could affect transcription of the *MSR1* gene. The missense change of Pro275Ala could affect the function of the MSR1 protein, because it changes a conserved residue in the first Gly-X-Y repeat of the collagenous domain of the protein. However, because all five sequence variants were associated with prostate cancer risk, and the risk haplotype "G (+) A C (−)" included all the risk alleles for each variant, it is difficult to dissect genetically which variant(s) are the most important changes. It is possible that one or more of these variants are associated with prostate cancer risk. Future functional analyses using various combinations of these variants may help to confirm these findings and provide insight into the function of each variant.

The results from this study and the study of *MSR1* rare mutations (Xu et al. 2002) suggest that rare mutations and common sequence variants of *MSR1* confer differential risks of prostate cancer. Although the rare *MSR1* mutations tend to impose relatively high risk of prostate cancer, common *MSR1* sequence variants within the same major genes tend to have a relatively low risk of prostate cancer. Similar observations were observed in two other major prostate cancer susceptibility genes, *HPC2/ELAC2* (MIM 605367) and *RNASEL* (MIM 180435) (Tavtigian et al. 2001; Carpten et al. 2002). For example, two rare mutations (E265X and M1I) of the ribonuclease L gene (RNASEL) were rare and observed mainly in families with HPC; another common sequence variant (Arg462Gln) was frequent in population and imposed intermediate risk to prostate cancer (Wang et al. 2002).

In summary, the significant differences in the allele and haplotype frequencies between patients with non-HPC and unaffected control subjects observed in this study suggest that common *MSR1* sequence variants are associated with prostate cancer risk in the general population. Together with evidence that the rare *MSR1* mutations are associated with increased prostate cancer susceptibility in patients with HPC and with non-HPC (Xu et al. 2002), the role of MSR1 and macrophages in prostate carcinogenesis is implicated. However, independent studies are extremely important to support these findings, given the complexity of prostate cancer.

Acknowledgments

The authors thank all the subjects who participated in this study. This work was partially supported by Public Health Service Specialized Program of Research Excellence CA58236, two grants from the Department of Defense (to W.B.I. and J.X.), and the Fund for Research and Progress in Urology, Johns Hopkins University.

Electronic-Database Information

Accession numbers and URLs for data presented herein are as follows:

Center for Human Genomics, http://www.wfubmc.edu/genomics/ (for authors' Web site)
GDA: Software for the Analysis of Discrete Genetic data, http://lewis.eeb.uconn.edu/lewishome/software.html
haplo.score, http://www.mayo.edu/statgen/software/
Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim/ (for prostate cancer [MIM 176807], *HPC2/ELAC2* [MIM 605367], *RNASEL* [MIM 180435], and *MSR1* [MIM 153622])

References

Carpten J, Nupponen N, Isaacs S, Sood R, Robbins C, Xu J, Faruque M, et al (2002) Germline mutations in the ribonuclease L gene in families showing linkage with HPC1. Nat Genet 30:181–184

De Marzo AM, Marchi VL, Epstein JI, Nelson WG (1999) Proliferative inflammatory atrophy of the prostate: impli-

cations for prostatic carcinogenesis. Am J Pathol 155:1985–1992

Emi M, Asaoka H, Matsumoto A, Itakura H, Kurihara Y, Wada Y, Kanamori H, Yazaki Y, Takahashi E, Lepert M, Lalouel JM, Kodama T, Mukai T (1993) Structure, organization, and chromosomal mapping of the human macrophage scavenger receptor gene. J Biol Chem 268:2120–2125

Excoffier L, Slatkin M (1995) Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. Mol Biol Evol 12:921–927

Kodama T, Freeman M, Rohrer L, Zabrecky J, Matsudaira P, Krieger M (1990) Type I macrophage scavenger receptor contains alpha-helical and collagen-like coiled coils. Nature 343:531–535

Lissbrant IF, Stattin P, Wikstrom P, Damber JE, Egevad L, Bergh A (2000) Tumor associated macrophages in human prostate cancer: relation to clinicopathological variables and survival. Int J Oncol 17:445–451

Nelson WG, De Marzo AM, DeWeese TL (2001) The molecular pathogenesis of prostate cancer: implications for prostate cancer prevention. Urology Suppl 57:39–45

Platt N, Gordon S (2001) Is the class A macrophage scavenger receptor (SR-A) multifunctional? The mouse's tale. J Clin Invest 108:649–654

Schaid DJ, Rowland CM, Tines DE, Jacobson RM, Poland GA (2002) Score tests for association between traits and haplotypes when linkage phase is ambigious. Am J Hum Genet 70:425–434

Shimura S, Yang G, Ebara S, Wheeler TM, Frolov A, Thompson TC (2000) Reduced infiltration of tumor-associated macrophages in human prostate cancer: association with cancer progression. Cancer Res 60:5857–5861

Tavtigian SV, Simard J, Teng DH, Abtin V, Baumgard M, Beck A, Camp NJ, et al (2001) A candidate prostate cancer susceptibility gene at chromosome 17p. Nat Genet 27:172–180

Wang L, McDonnell SK, Elkins DA, Slager SL, Christensen E, Marks AF, Cunningham JM. Peterson BJ, Jacobsen SJ, Cerhan JR, Blute ML, Schaid DJ, Thibodeau SN (2002) Analysis of the RNASEL gene in familial and sporadic prostate cancer. Am J Hum Genet 71:116–123

Weir BS (1996) Genetic data analysis II: methods for discrete population genetic data. Sinauer Association, Sunderland, MA

Xu J, Zheng SL, Komiya A, Mychaleckyj JC, Isaacs SD, Hu JJ, Sterling D, et al (2002) Germline mutations and sequence variants of the macrophage scavenger receptor 1 gene are associated with prostate cancer risk. Nat Genet 32:321–325

MR GG 2167

Evaluation of DLC1 as a prostate cancer susceptibility gene: mutation screen and association study

Siqun L. Zheng[a], Josyf C. Mychaleckyj[a], Gregory A. Hawkins[a], Sarah D. Isaacs[b], Kathy E. Wiley[b], Aubrey Turner[a], Bao-li. Chang[a], Chris. von Kap-Herr[a], John D. Carpten[c], Mark. Pettenati[a], Eugene R. Bleecker[a], Patrick C. Walsh[b], Jeffrey M. Trent[c], Deborah A. Meyers[a], William B. Isaacs[b*], Jianfeng. Xu[a]

[a]Center for Human Genomics, Wake Forest University School of Medicine, Winston-Salem, NC; [b]Department of Urology, Johns Hopkins Medical Institutions, Baltimore, MD; [c]National Human Genome Research Institute, National Institute of Health, Bethesda, MD

Running Title: Mutation screen and association of DLC1

Key words: prostate cancer, mutation screen, hereditary, association, DLC1, SNPs

*Address for correspondence:

William B. Isaacs, Ph.D.
Marburg 115
Johns Hopkins Hospital
600 N. Wolfe Street
Baltimore, MD 21287
Phone: (410) 955-2518
Fax: (410) 955-0833
Email: wisaacs@jhmi.edu

Abstract

A gene or genes on chromosome 8p22-23 have been implicated in prostate carcinogenesis by the observation of frequent deletions of this region in prostate cancer cells. More recently, two genetic linkage studies in hereditary prostate cancer (HPC) families suggest that germline variation in a gene in this region may influence prostate cancer susceptibility as well. DLC1 (deleted in liver cancer), a gene in this interval, has been proposed as a candidate tumor suppressor gene because of its homology (86% similarity) with rat p122 RhoGAP, which catalyzes the conversion of active GTP-bound rho complex to the inactive GDP-bound form, and thus suppresses Ras-mediated oncogenic transformation. A missense mutation and three intronic insertions/deletions in 126 primary colorectal tumors have been previously identified. However, there are no reports of DLC1 mutation screening in prostate tumors or in germ line DNA of prostate cancer patients. In this study, we report the results of the first mutation screen and association study of DLC1 in genomic DNA samples from hereditary and sporadic prostate cancer patients. The PCR products in the 5' UTR, all 14 exons, exon-intron junctions, and 3' UTR were directly sequenced in 159 HPC probands. Eight exonic nucleotide polymorphisms (SNPs) were identified, only one of which resulted in an amino acid change. Twenty-three other SNPs were identified in intronic regions. Seven informative SNPs that spanned the complete DLC1 gene were genotyped in an additional 249 sporadic cases and 222 unaffected controls. No significant difference in the allele and genotype frequencies were observed among HPC probands, sporadic cases, and unaffected controls. These results suggest that DLC1 is unlikely to play an important role in prostate cancer susceptibility.

1. Introduction

Prostate cancer is the most common cancer and the second leading cause of cancer death among men in the United States. In 2001, there will be an estimated 198,000 new prostate cancer cases diagnosed, accounting for over 30% of all cancers affecting men, with over 31,000 deaths annually resulting from this disease [1]. Although the etiology of prostate cancer is unknown, age, race/ethnicity, and family history are three well-established risk factors. Evidence for a prostate cancer susceptibility gene has been provided by segregation studies [2]. Several chromosomal regions that are likely to contain prostate cancer susceptibility genes have been identified in the past several years, including HPC1 at 1q24-25 [3], PCAP at 1q42-43 [4], HPCX at Xq27-28 [5], CAPB at 1p36 [6], HPC20 at 20q13 [7], and HPC2 at 17p11 [8].

Most recently, evidence for a prostate cancer susceptibility gene at 8p22-23 was provided by a genetic linkage analysis in 159 hereditary prostate cancer (HPC) families [9]. The prostate cancer linkage at this region was also observed in a recent genome-wide screen performed in 94 HPC families ascertained in the Seattle-based Prostate Cancer Genetic Research Study (PROGRESS) [10]. The likelihood of a prostate cancer susceptibility gene in this region is strengthened by the accumulated evidence that 8p is the site of the most frequent loss of heterozygosity (LOH) in prostate cancer tumors [11]. Several candidate tumor related genes reside in the 8p22-23 region, including N33 [12], macrophage-scavenger-receptor (MSR) [12], N-acetyltransferase (NAT)1 and NAT2 genes [13], FEZ1/LZTS1 [14], and DLC1 (deleted in liver cancer) [15, 16]. DLC1 is inferred to be a candidate tumor suppressor gene because of its homology (86% similarity) with rat p122 RhoGAP. The Rho family of proteins is a subfamily of the Ras

small GTP binding superfamily, and the product of the RhoGAP gene can catalyze the conversion of active GTP-bound rho complex to the inactive GDP-bound form, thus suppressing Ras-mediated oncogenic transformation. Screening for point mutations of DLC1 in colorectal and ovarian tumors has been reported. A missense mutation and three intronic insertions/deletions were identified in 126 primary colorectal tumors [16]. However, to our knowledge, no mutation screening for DLC1 in prostate cancer tumors or genomic DNA of prostate cancer patients has been reported.

To test the hypothesis that DLC1 is a prostate cancer susceptibility gene, we performed two types of analyses. We first screened the 5' UTR, all 14 exons, exon-intron junctions, and the 3' UTR of DLC1 for mutations and sequence variants in 159 HPC probands. We then tested for an association between prostate cancer and DLC1 by genotyping 7 informative or highly polymorphic SNPs that provide complete coverage of the DLC1 gene, in an additional 249 sporadic prostate cancer patients and 222 unaffected controls.

2. Materials and Methods

2.1. Subjects

A detailed description of the study sample was presented elsewhere [9]. Briefly, a total of 159 HPC probands were ascertained at the Brady Urology Institute of Johns Hopkins Hospital (Baltimore, MD), through referrals, medical records of patients seen at Johns Hopkins Hospital for treatment of prostate cancer, and respondents to various lay publications describing our studies. All of the probands had at least two first degree relatives affected with prostate cancer. The diagnosis of prostate cancer was verified by medical records. The mean age at prostate cancer diagnosis for these probands was 61

years. Among the probands, 133 (84%) are Caucasian and 14 (8.8%) are African American.

All 245 unrelated prostate cancer cases were recruited from patients who underwent treatment for prostate cancer at the John Hopkins Hospital and did not have first-degree relatives affected with prostate cancer. The diagnosis of prostate cancer for all these subjects was confirmed by pathology reports. Preoperative prostate specific antigen (PSA) levels, Gleason score, and pathological stages were available for 202, 240, and 241 cases, respectively. Mean age at diagnosis for these cases was 58.7 years. Over 93% of the cases are Caucasian, and 3.2% are African American.

Two hundred and twenty-two non-prostate cancer controls were selected from men participating in screening programs for prostate cancer. By applying the exclusion criteria of abnormal digital rectal examination (DRE) and abnormal PSA level (i.e., >=4 ng/ml), 211 were eligible for the study. The mean age at examination was 58 years. Over 86% of the eligible controls were Caucasian and 7.1% were African American. About 5.6% of the eligible controls have brothers or father affected with prostate cancer. The affection status of relatives was obtained by interview of the probands.

2.2. Sequencing methods and SNPs genotyping

Wilson et al. [16] previously described the genomic organization of DLC1 (13 exons) by sequencing a combination of long-range PCR products spanning introns and a PAC clone encompassing the gene. However, they cautioned that additional exons might exist, as the sequence containing the initial methionine was not amplified in their study, and they were not able to publish a size for their putative exon 1. To define the precise

gene organization, we took advantage of the public reference human genome sequence. The full length DLC1 mRNA (GenBank accession NM_006094) was optimally aligned with genome contig NT_008161 gi:14749165, using the Smith-Waterman alignment program swat (gap extension penalty = 0) from Phrep suite (P. Green, unpublished), and then manually corrected for consensus donor/acceptor splice site location. Table 1 shows the results of this bioinformatics analysis, including all 14 exons, and 13 introns. The gene orgnization of DLC1 is shown in Figure 1. The geneplot program, written in Perl, was used to map, draw, and annotate the gene structure (J. Mychaleckyj, unpublished).

We directly sequenced the PCR products of the 5' UTR, all 14 exons, exon-intron junctions, and the 3' UTR of DLC1 in 159 HPC probands. Table 2 lists the primers used to amplify the PCR products, the sizes of amplified PCR fragments, and the annealing temperatures for each pair of primers. All PCR reactions were performed in a 30ul volume consisting of 10ng genomic DNA, 0.2uM of each primer, 0.2mM of each dNTP, 1.5 mM $MgCl_2$, 20 mMTris-HCl, 50mM KCl, and 0.5 u Taq polymerase (Life Technologies, Inc.). PCR cycling conditions were as follows: 94^0C for 4 minutes; followed by 30 cycles of 94^0C for 30 seconds, specified annealing temperature for 30 seconds, and 72^0C for 30 seconds; with a final extension of 72^0C for 6 minutes. All PCR products were purified using the QuickStep ™PCR purification Kit (Edge BioSystems, Gaithersburg, MD) to remove dNTPs and excess primers. All sequencing reactions were performed using dye-terminator chemistry (BigDye, ABI, Foster City,CA) and then precipitated using 63+/-5% ethanol. Samples were loaded onto an ABI 3700 DNA Analyzer after adding 8ul of formamide. SNPs were identified using Sequencher™ software version 4.0.5 (Gene Codes Corporation). Seven informative SNPs were

genotyped in an additional 249 sporadic prostate cancer cases and 222 unaffected controls using the same sequencing method.

2.3. Statistical methods

Hardy-Weinberg Equilibrium (HWE) tests for all SNPs, and linkage disequilibrium (LD) tests for all pairs of SNPs, were performed using the GDA computer program [17]. The HWE tests were based on exact tests, where a large number of the possible arrays are generated by permuting the alleles among genotypes, and the proportion of these permuted genotypic arrays that have a smaller conditional probability than the original data is calculated. The LD tests were based on an exact test assuming multinomial probability of the multilocus genotype, conditional on the single-locus genotype [18]. A Monte Carlo simulation was used to assess the significance, by permuting the single-locus genotypes among individuals in the sample to simulate the null distribution. The empirical p-values of both the HWE and LD tests were based on 10,000 replicate samples.

Tests for associations between the SNPs and prostate cancer were performed by comparing allele and genotype frequencies between cases and controls for each SNP. Allele frequencies were estimated by direct count. The hypotheses of differences in allele frequencies between cases and controls were tested using standard contingency χ^2 tests, and P-values were determined via χ^2 approximation. The hypotheses of differences in genotype frequencies between cases and controls were tested using ANOVA. Unconditional logistic regression was used to test for the differences of genotype frequencies between cases and controls, adjusted for potential confounders such as age.

3. Results

3.1. SNP identification

A total of 31 SNPs were identified in the sequenced region of DLC1 in 159 HPC probands. The location and the frequency of each SNP in whites and blacks are presented in Table 3 and Figure 1. Eight of these SNPs are in the exons. However, only one SNP (WF100-011) results in an amino acid change from valine to methionine at codon 354. The remaining 23 SNPs are in the intronic regions. Eleven of the 31 SNPs are polymorphic, with a less frequent allele >5%. We only observed two [WF100-009 (Ala170Ala) and WF100-011 (Val354Met)] of the 10 exonic SNPs identified by Wilson et al., [16].

3.2. Association between prostate cancer susceptibility and the SNPs

The seven frequent SNPs, spanning the entire DLC1 gene, were further genotyped in an additional 249 sporadic prostate cancer cases and 222 unaffected controls. To decrease the potential impact of population stratification, all of the following analyses were limited to Caucasians. All of the SNPs were in HWE in the HPC probands, sporadic cases, and controls. Pair-wise LD tests for all SNPs were also performed. SNP WF100-001 was in linkage equilibrium with each of the other SNPs (all P-values >0.05). This SNP is at least 17 kb away from the rest of the SNPs. The rest of the SNPs were in strong LD (all P-values $<10^{-5}$).

There was no statistical difference in the allele frequencies between hereditary prostate cancer patients, sporadic prostate cancer, and unaffected controls (Table 4). The

largest difference in allele frequency between these groups was observed at the SNP WF100-018. The frequency of allele '2' of this SNP was 0.63 in the HPC cases, 0.59 in the sporadic cases, and 0.55 in the controls. The frequency of the valine to methionine change associated with SNP WF100-011 was similar in HPC cases (0.48), sporadic cases (0.49), and unaffected controls (0.47). When the genotypic frequencies of these SNPs were compared between the cases and controls, no significant difference was observed (data not shown), with or without an adjustment for age.

3.3. Association between characteristics of prostate cancer and the SNPs

Relationships between the seven frequent SNPs and either Gleason scores or pathological stages in sporadic prostate cancer cases were also examined. No statistically significant difference in the genotypic frequencies of these SNPs was found between the groups with low (≤6) or high (≥7) Gleason scores, or between the groups with disease confined to the prostate versus non-localized disease (data not shown).

4. Discussion

Evidence for a prostate cancer gene (or genes) on 8p has been observed in linkage studies and loss of heterozygosity (LOH) studies. Three independent linkage studies have provided evidence for a prostate cancer susceptibility gene(s) on 8p22-23. The first was from a genome-wide screen in 66 prostate cancer families reported by our group. A two point parametric LOD of 0.7 at D8S550, a multipoint LOD assuming heterogeneity (HLOD) of 0.81 (p=0.05), and a multipoint non-parametric linkage (NPL) score of 2.02 (p=0.02) was observed at 8p22-23 [3]. The evidence for linkage at this region was

stronger in a recent study, after we genotyped additional markers in the region and included an additional 93 families (total 159 HPC families) [9]. In the complete set of families, evidence for prostate cancer linkage was found at 8p22-23, with a peak HLOD of 1.84 (p=0.004) at D8S1130. In the 79 families with an average age of diagnosis over 65, an allele sharing LOD score of 2.64 (p=0.0005) at the region was observed. Six markers spanning a distance of 10 cM had LOD scores >2.0, including DLC1 (near D8S1106). The second study was a recent genome-wide screen performed in 94 HPC families ascertained in the Seattle-based Prostate Cancer Genetic Research Study (PROGRESS) [10]. This study reported a 19 cM positive linkage region on 8p22-23, with a maximum multipoint NPL score of 2.02 (p=0.026) at D8S1106 in 44 pedigrees with late age of onset (>=66 years). The third supporting evidence was from a linkage study in 254 families with siblings affected with prostate cancer [19]. The LOD score was 1.92 (p=0.003) at 8p22-23 when number of affected siblings was included as a covariate. The possibility of a prostate cancer susceptibility gene (or genes) on 8p is increased by the consistent finding from many LOH and allelic imbalance (AI) studies that 8p is the most frequently deleted region in prostate cancer. Latil and Lidereau [20] reviewed over 30 published prostate cancer LOH studies and found that among 800 tumors examined, 66% had LOH at 8p.

DLC1 is a candidate gene for prostate cancer susceptibility because of its chromosomal location and potential tumor suppressor function. However, the results from our study suggest that DLC1 is unlikely to be an important prostate cancer susceptibility gene. This conclusion is based on the following two observations. First, no potentially important mutation, such as a protein-truncating mutation, was found after

screening for mutations of DLC1 in germ line DNA samples of 159 HPC probands. Although one missense mutation (Val354Met) was found, it is unlikely to be important because both valine and methionine are neutral and hydrophobic, and more importantly, there is no difference in the frequency of the Val354Met carrier rate between HPC probands, sporadic cases, and unaffected controls. Second, no statistically significant difference in the allele, genotype, and haplotype frequencies of any of the SNPs were found after genotyping 7 frequent SNPs in 159 HPC probands, 249 sporadic cases, and 222 unaffected controls. The 7 SNPs span ~ 47 kb and provide good coverage of the DLC1 gene (~48 kb). Except for the first SNP, which was in linkage equilibrium with the other SNPs, the remaining SNPs are in strong linkage disequilibrium with each other. Thus the information obtained from these SNPs also represents other potential sequence variants in the gene.

Caution should be taken when interpreting and generalizing the results from our study. The power to detect an association is limited in our study. Although our study has reasonable power to detect an association between prostate cancer and sequence variants that have a large effect and high frequency, this study has limited power to detect an association between prostate cancer and sequence variants that have a small effect and low frequency. For example, when a sequence variant is present in 10% of the controls, the power to detect a sequence variant that confers OR of 2.5 and 1.5, is 95% and 31%, respectively. There is also a potential for misclassification in our control subjects, which may come from two sources. Unaffected controls in our study were from a prostate cancer screening population, and thus may be likely to represent a high-risk group (self selected). This potential bias, however, is unlikely to be significant, because all controls

were carefully examined and found to have normal digital rectal examination (DRE) and PSA results. Furthermore, the majority of the controls do not report a positive family history. Based on the collection of detailed family history information from each of the controls, there were only 6 controls who reported a positive family history (defined as an affected father and/or brothers) among 182 Caucasian controls. We performed additional analyses excluding the six individuals and the results were similar. Potential misclassification may also come from the fact that some controls are still young and they may become affected at a later date. We are aware of this potential bias and performed analyses adjusting for age, which may not remove, but may alleviate the problem. We also performed an additional analysis, which included only older control individuals (>60 years). No statistically significant difference in the allelic and genotypic frequencies of the SNPs was observed between cases and this subset of controls (data not shown). Another important caveat of the study is the possibility that sequence variants in the promoter region of DLC1 may affect the expression of this gene.

Our results are consistent with the findings of a mutation screening of DLC1 in colorectal and ovarian primary tumor and cell lines by Wilson et al [16]. By studying 104 primary colorectal and 26 primary ovarian tumors, as well as 22 colorectal and 7 ovarian cancer cell lines, using a combination of SSCP and direct sequencing, this group identified only one missense change in a primary colorectal tumor and two missense changes in colorectal cell lines. They concluded that DLC1 is not the target of 8p LOH in colorectal or ovarian tumors, based on the rarity of exonic missense mutations and the absence of protein-truncating mutations. It is worth noting that we did not find these sequence variants in the germline DNA samples of the159 probands.

Results of our study suggest that other important candidate genes in the 8p22-23 regions should be analyzed. Furthermore, several recent studies suggest that 8p genomic sequence is characterized by highly repetitive gene family members, including Alu, LINE, LTR, and olfactory receptor, which in turn are responsible for the formation of recurrent chromosomal changes [21]. In particular, a submicroscopic inversion at 8p23 is observed 26% in normal Caucasians [21]. Studies that utilize multiple approaches, such as cytogenetics, molecular, association, and bioinformatics, most likely will be required to identify prostate cancer susceptibility genes on 8p.

Acknowledgements

The authors thank all the study subjects who participated in this study. This work was partially supported by PHS SPORE CA58236 and two grants from the Department of Defense to (W.B.I and J.X), and Ella & Georg Ehrnrooth Foundation.

References

[1] R.T. Greenlee, T. Murray, S. Bolden, P.A. Wingo, Cancer statistics, 2000, CA Cancer J Clin. 50 (2000) 7-33.

[2] W.B. Isaacs, J. Xu, P.C. Walsh, Hereditary prostate cancer, in: L.W.K. Chung, W.B. Isaacs, J.W. Simons (Eds.), Prostate cancer, biology, genetics, and the new therapeutics, Humana Press, New Jersey, 2001, pp. 13-28.

[3] J.R. Smith, D. Freije, J.D. Carpten, H. Gronberg, J. Xu, S.D. Isaacs, M.J. Brownstein, G.S. Bova, H. Guo, P. Bujnovszky, D.R. Nusskern, J.E. Damber, A. Bergh, M. Emanuelsson, O.P. Kallioniemi, J. Walker-Daniels, J.E. Bailey-Wilson, T.H. Beaty, D.A. Meyers, P.C. Walsh, F.S. Collins, J.M. Trent, W.B. Isaacs, Major susceptibility locus for prostate cancer on chromosome 1 suggested by a genome-wide search, Science 274 (1996) 1371-1374.

[4] P. Berthon, A. Valeri, A. Cohen-Akenine, E. Drelon, T. Paiss, G. Wohr, A. Latil, P. Millasseau, I. Mellah, N. Cohen, H. Blanche, C. Bellane-Chantelot, F. Demenais, P. Teillac, A. Le Duc, R. de Petriconi, R. Hautmann, I. Chumakov, L. Bachner, N.J.

Maitland, R. Lidereau, W. Vogel, G. Fournier, P. Mangin, O. Cussenot O, Predisposing gene for early-onset prostate cancer, localized on chromosome 1q42.2-43, Am. J. Hum. Genet. 62 (1998) 1416-1424.

[5] J Xu, D. Meyers, D. Freije, S. Isaacs, K. Wiley, D. Nusskern, C. Ewing, E. Wilkens, P. Bujnovszky, G.S. Bova, P. Walsh, W. Isaacs, J. Schleutker, M. Matikainen, T. Tammela, T. Visakorpi, O.P. Kallioniemi, R. Berry, D. Schaid, A. French, S. McDonnell, J. Schroeder, M. Blute, S. Thibodeau, J. Trent J, Evidence for a prostate cancer susceptibility locus on the X chromosome, Nat. Genet. 20 (1998) 175-179.

[6] M. Gibbs, J.L. Stanford, R.A. McIndoe, G.P. Jarvik, S. Kolb, E.L. Goode, L. Chakrabarti, E.F. Schuster, V.A. Buckley, E.L. Miller, S. Brandzel, S. Li, L. Hood, E.A. Ostrander, Evidence for a rare prostate cancer-susceptibility locus at chromosome 1p36, Am. J. Hum. Genet. 64 (1999) 776-787.

[7] R. Berry, J.J. Schroeder, A.J. French, S.K. McDonnell, B.J. Peterson, J.M. Cunningham, S.N. Thibodeau, D.J. Schaid, Evidence for a prostate cancer-susceptibility locus on chromosome 20, Am. J. Hum. Genet. 67 (2000) 82-91.

[8] S.V. Tavtigian, J. Simard, D.H. Teng, V. Abtin, M. Baumgard, A. Beck, N.J. Camp, A.R. Carillo, Y. Chen, P. Dayananth, M. Desrochers, M. Dumont, J.M. Farnham, D. Frank, C. Frye, S. Ghaffari, J.S. Gupte, R. Hu, D. Iliev, T. Janecki, E.N. Kort, K.E. Laity,

A. Leavitt, G. Leblanc, et al., A candidate prostate cancer susceptibility gene at chromosome 17p, Nat. Genet. 27 (2001) 172-80.

[9] J. Xu, S.L. Zheng, B. Chang, S.D. Isaacs, K. Wiley, G.A. Hawkin, E.R. Bleecker, P.C. Walsh, J.M. Trent, D.A. Meyers, W.I. Isaacs, Linkage and association studies of prostate cancer susceptibility gene on 8p22-23, Am. J. Hum. Genet. 69 (2001) 341-350.

[10] M. Gibbs, J.L. Stanford, G.P. Jarvik, M. Janer, M. Badzioch, M.A. Peters, E.L> Goode, S. Kolb, L. Chakrabarti, M. Shook, R. Basom, E.A. Ostrander, L. Hood, A genomic scan of families with prostate cancer identifies multiple regions of interest, Am. J. Hum. Genet. 67 (2000) 100-9.

[11] R. Bookstein, Tumor suppressor genes in prostate cancer, in: L.W.K. Chung, W.B. Isaacs, J.W. Simons (Eds.), Prostate cancer, biology, genetics, and the new therapeutics, Humana Press, New Jersey, 2001, pp. 61-94.

[12] G.S. Bova, D. MacGrogan, A. Levy, S.S. Pin, R. Bookstein, W.B. Isaacs, Physical mapping of chromosome 8p22 markers and their homozygous deletion in a metastatic prostate cancer, Genomics 35 (1996) 46-54.

[13] C.Y. Wang, M. Debiec-Rychter, H.A. Schut, P. Morse, R.F. Jones, C. Archet, C.M> King, G.P. Haas, N-Acetyltransferase expression and DNA binding of N-

hydroxyheterocyclic amines in human prostate epithelium, Carcinogenesis 20 (1995) 1591-5.

[14] H. Ishii, R. Baffa, S.I. Numata, Y. Murakumo, S. Rattan, H. Inoue, M. Mori, V. Fidanza, H. Alder, C.M. Croce, The FEZ1 gene at chromosome 8p22 encodes a leucine-zipper protein, and its expression is altered in multiple human tumors, Proc. Natl. Acad. Sci. U. S. A. 96 (1999) 3928-33.

[15] B.Z. Yuan, Y. Yang, C.L. Keck-Waggoner, D.B. Zimonjic, S.S. Thorgeirsson, N.C. Popescu, Assignment and cloning of mouse Arhgap7 to chromosome 8A4-B2, a conserved syntenic region of human chromosome 8p22-->p21, Cytogenet. Cell Genet. 87 (1999) 18.

[16] P.J. Wilson, E. McGlinn, A. Marsh, T. Evans, J. Arnold, K. Wright, K. Biden, J. Young, B. Wainwright, C. Wicking, G. Chenevix-Trench, Sequence variants of DLC1 in colorectal and ovarian tumours, Hum. Mutat. 15 (2000) 156-65.

[17] B.S. Weir, Genetic Data Analysis II: Methods for Discrete Population Genetic Data, Sinauer Association, Inc. Publishers, Sinauer, Sunderland, MA, (1996)

[18] D. Zaykin, L. Zhivotovsky, B.S. Weir, Exact tests for association between alleles at arbitrary numbers of loci, Genetica, 96 (1995) 169-78.

[19] K.A. Goddard, J.S. Witte, B.K. Suarez, W.J. Catalona, J.M. Olson, Model-free linkage analysis with covariates confirms linkage of prostate cancer to chromosomes 1 and 4, Am. J. Hum. Genet. 68 (2001) 1197-206.

[20] A. Latil, R. Lidereau, Genetic aspects of prostate cancer, Virchows Arch 432 (1998) 389-406.

[21] S. Giglio, K.W. Broman, N. Matsumoto, V. Calvari, G. Gimelli, T. Neumann, H. Ohashi, L. Voullaire, D. Larizza, R. Giorda, J.L. Weber, D.H. Ledbetter, O. Zuffardi, Olfactory receptor-gene clusters, genomic-inversion polymorphisms, and common chromosome rearrangements, Am. J. Hum. Genet. 68 (2001) 874-83.

Table 1. Predicted Gene Organization for DLC1

Gene Element	Size (bp)	5' Sequence	3' Sequence
Exon 1	333	GCCCGAGCGAGGGC	CATGATCCTAACAC
Intron 1	17289	gtaagctagacttg	ctgttctgttctag
Exon 2	72	AAATTGAAGCCAAG	ACAGCTTTATGAAG
Intron 2	4762	gtaagctgggaatg	tatttttctcatag
Exon 3	82	ATTTCCTGTTCCCC	GAGGCTCTATGCAG
Intron 3	7886	gtaaatggacactt	tctttgctctatag
Exon 4	64	GCGTCTAAATACTT	CTCATCGGAAACGA
Intron 4	2017	gtgagtatacaaat	ttctgtctttgcag
Exon 5	1424	AGTGACGATTCAGA	ACCAGGTCCAACAG
Intron 5	771	gtaagaacttttct	ctcttctctcgcag
Exon 6	177	GCACCGACTGAGAT	CATGGTTTTAGCTG
Intron 6	3159	gtaagagtttaaat	gtgttctttaacag
Exon 7	160	GGCCGTGCCCAAGT	ATTGTTTGGATCAG
Intron 7	128	gtgagagcgctgcc	tgctcttccgacag
Exon 8	199	GTTGGGCTCTTCAG	ACAGATCTACCAAT
Intron 8	1932	gtgagtgtcctttg	ctggtgtcgtgcag
Exon 9	214	ATGTGCCCAAGGAC	AATTCCTCTCCCAG
Intron 9	1179	gtacgggctgcatg	tttcctctacctag
Exon 10	115	GGTAATGCAAAGAA	AGAAGCTTTTCCAG
Intron 10	848	gtaaggaattgaga	catgttccgcacag
Exon 11	219	GTTCCCGAGGAAAT	TGTCCTATAAGAAG
Intron 11	1545	gtaaggcttcaccc	catcctctttccag
Exon 12	218	GTGAGCGAAGGACC	TACGTTGTTTTAAG
Intron 12	2023	gtgagcgcttccca	ttttgtcgcgacag
Exon 13	174	AACCTGGAGGACTA	AGAGTTGACTTAAG
Intron 13	358	gtatgttctgattc	cttttctctgatag
Exon 14	339	GGGCCACATGCCAG	ATGTAAGACTGTGC

Gene element sequences and size estimates are derived from reference human genomic sequence, in contig NT_008161 gi:14749165.

Table 2. Primers used for mutational analysis and SNPs identification

PCR ID	Forward primer	Reverse primer	PCR Product (bp)	Annealing temperature
1	AGCAAGGATGCGTTGAGGAC	GGAGACTCTGCAGAAAGCG	305	55
2	CAGTCATGTTTTGCTTGTATC	AGGGATTATGATTATGCCAAC	222	55
3	CCACATTAAAGCCGAAGC	CTAGGACAGCGTGGATG	501	62
4	GGTGTTTATCCACTCATTAG	TAGCTGATGTGATCAGGTC	361	60
5	GCATATTATGAGCAGCCTCC	CAAGTTGCTGGAGGAGCAAACG	411	63
6	AGGTGTCTTCCGTCCGCAG	TTCTCCTCATCCATCCCCTC	352	55
7	GATGGAGAGCCTGAAGCTC	GTTCACAGAGCCGTTATTCC	502	63
8	CAGAGGACACGGTGTTCTAC	CAGAACAAGCACAGGCAATG	657	65
9	CTACAGTCCTGGCTACTCAG	CAGAGAGTGATTGCCATAGG	500	57
10	GTTTGCCTGCAGAATTGGAC	CAACACTAAGTGTGGGGTAC	358	65
11	CAGCTTCTTCTAGTTGTGTGG	CAGAGCGAGACTCCATCTC	388	65
12	GTGCACTTTCTACAAGTGC	GGTCTAACCTTCTTGCTTG	459	61
13	CTGAGATACTCTGGTGACC	CAAGCCTAAGATTTTGTACC	349	62
14	GACCAGGCCATTGTTCCTAC	CTCTCCCAAGTGTTACACAG	531	58
15	TTCACTTTCTGGCTTGCTC	GTAAGATCAGTGACATATAGGC	527	62
16	CGTCACCACACTGCAAATAG	GAATTCATGGCCGAGCTAC	509	67
17	CTCGGCCATGAATTCAAAGG	CGCATACACTGATATCACAAGAG	593	62

Table 3. SNPs of DLC1 identified in 159 HPC probands

Predicted Location	SNP Identifier	Nucleotide Change†	Codon Change	Frequency	
				Whites (290)	Blacks (n=28)
Exon 1	WF100-001	-29A>T		0.28	0.28
Intron 1	WF100-002	133T>C		0.28	0.33
Intron 1	WF100-003	17292A>G		0	<0.05
Intron 2	WF100-004	17415A>G		0.52	0.32
Intron 4	WF100-005	30210A>G		0.50	0.15
Intron 4	WF100-006	30275T>C		<0.01	<0.05
Intron 4	WF100-007	30341T>C		<0.01	0
Exon 5	WF100-008	32344A>G	Pro130Pro	0.02	<0.0.5
Exon 5	WF100-009	32464G>T	Ala170Ala	0.27	0.25
Exon 5	WF100-010	32842C>T	Ser296Ser	0	0.07
Exon 5	WF100-011	33014G>A	Val354Val	0.49	0.50
Intron 5	WF100-012	34357T>C		<0.01	0
Intron 6	WF100-013	34628T>C		0	<0.05
Intron 6	WF100-014	34652T>C		0	<0.05
Intron 6	WF100-015	37676G>A		<0.01	0
Intron 7	WF100-016	37943G>T		<0.01	0
Exon 8	WF100-017	38202G>A	Ser730Ser	0.06	0.09
Intron 8	WF100-018	38265G>A		0.64	0.60
Intron 8	WF100-019	38311T>C		0	<0.05
Exon 9	WF100-020	40179C>A	Arg745Arg	0.07	0.10
Intron 9	WF100-021	41485T>G		<0.01	<0.05
Intron 9	WF100-022	41496T>G		0	0.07
Intron 9	WF100-023	41498G>A		<0.01	0
Intron 9	WF100-024	41543T>C		0.57	0.75
Intron 10	WF100-025	41714A>G		<0.01	0.07
Exon 11	WF100-026	42656G>A	Leu895Leu	<0.01	0

Intron 11	WF100-027	42787-insC	0	0.07
Intron 11	WF100-028	44201T>C	0	0.07
Intron 13	WF100-029	46698T>A	<0.01	0
Intron 13	WF100-030	46802A>G	0.43	0.32
Intron 13	WF100-031	46868G>C	0	0.07

† All positions are quoted for reference genomic sequence contig, accession NT_008161, gi:14749165. The first base of the initial methionine ATG codon is designated +1, and the base immediately 5' to this is –1, as per convention.

Table 4. Allelic frequencies of DLC1 SNPs in cases and controls (Whites)

SNPs	Allele	HPC	Sporadic	Control	P-values[1]	P-values[2]	P-values[3]
WF100-001	A	0.74	0.7	0.66	0.16	0.27	0.15
WF100-004	A	0.49	0.51	0.5	0.83	0.69	0.86
WF100-005	A	0.51	0.51	0.5	0.89	0.86	0.85
WF100-011	A	0.48	0.49	0.47	0.74	0.53	0.55
WF100-018	A	0.63	0.59	0.55	0.06	0.2	0.07
WF100-024	T	0.45	0.42	0.45	0.92	0.32	0.44
WF100-030	A	0.55	0.56	0.52	0.42	0.29	0.27

Note: χ^2 tests: [1]HPC probands vs. controls, [2]Sporadic cases vs. controls, [3]combined cases vs. controls

Figure legend

Figure 1. Gene orgnization of DLC1. The full length DLC1 mRNA (GenBank accession NM_006094) was optimally aligned with genome contig NT_008161 gi:14749165, using the Smith-Waterman alignment program swat (gap extension penalty = 0) from Phrep suite (P. Green, unpublished), and then manually corrected for consensus donor/acceptor splice site location. The DLC1 gene has 14 exons and 13 introns. SNPs identified in the 159 HPC probands were also indicated.